

New Frontiers in Grey Literature

Sponsors

BIOSIS Publisher of Biological Abstracts and Zoological Record

JST Japan Science and Technology Corporation

MCB University Press
Networked Publisher in
Management, Engineering and
Library Science

NASA National Aeronautics
and Space Administration

NLE National Library of
Education, United States
Department of Education

Program Committee

BIOSIS Publisher of Biological Abstracts and Zoological Record

EAGLE European Association
for Grey Literature Exploitation

FID International Federation for
Information and Documentation

GreyNet
Grey Literature Network Service

IFLA International Federation
of Library Associations and
Institutions

JST Japan Science and
Technology Corporation

MCB University Press
Networked Publisher in
Management, Engineering and
Library Science

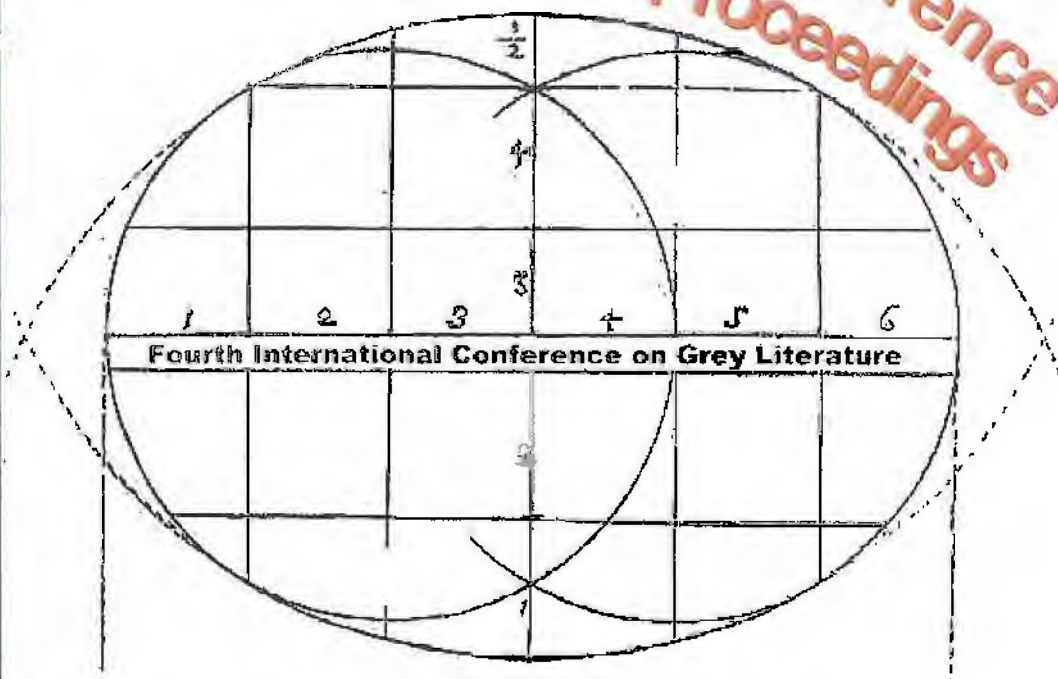
NASA National Aeronautics
and Space Administration

NLE National Library of
Education, United States
Department of Education

GL'99

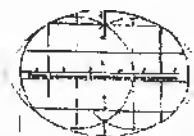
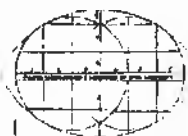
Kellogg Conference Center
Washington D.C., USA
4-5 October 1999

Conference
Proceedings



Program and Conference Bureau:

GreyNet, Grey Literature Network Service
Koninginneweg 201, 1075 CR Amsterdam, The Netherlands
Tel/ Fax: 31-20-671.1818 - GreyNet@inter.nl.net
<http://www.konbib.nl/infolev/greynet>



New Frontiers in Grey Literature
FOURTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

● Reports ● Proceedings ● Dissertations ● Bibliographies ● Manuals
● CD-ROM ● Databases ● Bulletin Boards ● WWW ● Email ● Internet ● Etc.

Kellogg Conference Center

Washington D.C., USA

4-5 October 1999



GL'99 Program and Conference Bureau

GreyNet, Grey Literature Network Service

Koninginneweg 201, 1075 CR Amsterdam

The Netherlands

Phone/Fax : 31-20-671.1818

Email : GreyNet@inter.nl.net

<http://www.konbib.nl/infolev/greynet/GL'99.htm>

Conference Proceedings on Grey Literature : GL-conference series /
compiled by D.J. Farace ; with the assistance of J. Frantzen. - Amsterdam :
GreyNet, Grey Literature Network Service. - ISSN 1386-2316

No. 1: GL'93 Conference Proceedings - 416 p. - 1994. - ISBN 90-74854-03-6 : € 55

No. 2: GL'95 Conference Proceedings - 266 p. - 1996. - ISBN 90-74854-08-7 : € 55

No. 3: GL'97 Conference Proceedings - 294 p. - 1998. - ISBN 90-74854-17-6 : € 55

No. 4: GL'99 Conference Proceedings - 322 p. - 2000. - ISBN 90-74854-24-9 : € 55

CIP

GL'99 Conference Proceedings

Fourth International Conference on Grey Literature : GL'99 Conference Proceedings :
New Frontiers in Grey Literature, Washington D.C., 4-5 October 1999 / compiled
by D.J. Farace ; with the assistance of J. Frantzen. - Amsterdam : GreyNet,
Grey Literature Network Service, January 2000. - XII, 322 p. ; 30 cm. - Index. -
(GL-Conference series, ISSN 1386-2316 ; No. 4)

*Contains the full text of the twenty-eight papers presented during the Opening, Plenary and
Breakout Sessions*

Acknowledgements

On behalf of the Sponsors and Program Committee Delegates, it is my distinct pleasure to welcome you, the authors and participants to this, the Fourth International Conference on Grey Literature. You have travelled here, to the capital of the United States of America, from more than twenty countries worldwide. Together, you represent more than one hundred organisations in the various sectors and levels of government, academics, business and industry. We have been brought together by the common theme "New Frontiers in Grey Literature", to discuss new worlds of topics, formats, uses and applications of Grey Literature - both in print and electronic formats, not controlled by commercial publishers. Nevertheless, we hold this information to be valuable - responsible often in the first place for that which later only becomes recognized in small measure. Our frontiers cross time and space. Temporal, because we are on the eve of a new century, a new millennium. Spatial, belonging to geographical, regional, as well as, virtual space and locators.

We are here not as one, but as many and varied; however, we share a common vision, a noble mission - not as pioneers but instead as that which we are:

- Librarian and Documentalists,
- Professors and Lecturers,
- Editors and Publishers,
- Managers and Policy Makers,
- Project Coordinators & Database Administrators,
- Journalists and Business Analysts,
- Researchers and Information Specialists,
- Etcetera.

No, we are not one, however we are truly united, absolutely diversified. We here, span fields and cross the disciplines of agriculture, rural development, aerospace, applied sciences, military and defence, physics, social sciences, economic, computer science, transportation studies, medicine, education, management and engineering, political science and international affairs, biology, zoology, energy, environment, and last but not least, library and information science.

Yes, you are all truly welcome to GL'99. Your one and real mission - not impossible, but instead inspired - to demonstrate at this international forum, the results of 49 authors and co-authors months of research and writing. Bridging theoretical constructs and academic terminology with empirical facts, models, and paradigms.

At this fourth international conference, perhaps a bit more than at the past three in this series, we stand at the crossroads, not on the periphery, but at the very heart of this grey matter. Now, ready more than ever to demonstrate that grey literature is not merely the document in electronic and print format, but is that which encompasses the very process of communication responsible for knowledge transfer - encoded and deciphered.

Dr. Dominic J. Farace
Program and Conference Director

Conference Sponsors



BIOSIS
Publisher of Biological Abstracts and Zoological Record



JST
Japan Science and Technology Corporation



MCB University Press
Networked Publisher in Management, Engineering, and
Library Science

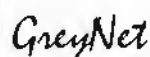


NASA
National Aeronautics and Space Administration



NLE
National Library of Education,
United States Department of Education

Program & Conference Bureau



Grey Literature Network Service
Koninginneweg 201
1075 CR Amsterdam, The Netherlands

Tel/Fax: 31-671.1818
Email: GreyNet@inter.nl.net
URL: <http://www.konbib.nl/infolev/greynet/GL'99.htm>

Program Committee

BIOSIS	Bruce H. Kiesel, Director IAD Publisher of Biological Abstracts and Zoological Record
EAGLE	Andrew Smith, Chairman Technical Committee European Association for Grey Literature Exploitation
FID	Martha B. Stone, President International Federation for Information and Documentation
GREYNET	Dominic J. Farace, Head Grey Literature Network Service
IFLA	Graham P. Cornish, Director UAP International Federation of Library Associations and Institutions
JST	Masayuki Sato, Director Washington Office Japan Science and Technology Corporation
MCB	Eileen Breen, Managing Editor MCB University Press
NASA	Carolyn E. Floyd, Information System and Services Division National Aeronautics and Space Administration
NLE	E. Stephen Hunt, Director for Planning and Policy National Library of Education; U.S. Department of Education

Conference Program

Opening Session

Chairperson: Dominic J. Farace, GreyNet
Grey Literature Network Service

Session One

Global Assessment of Grey Literature
A Brave New World of Topics, Formats, and Uses

Chairperson: E. Stephen Hunt, NLE
National Library of Education, United States Department of education

Session Two

Publishing and Archiving Electronic Grey Literature
From Production to Full-Text Storage, Retrieval and Distribution

Chairperson: Eileen Breen, MCB University Press
Networked Publisher in Management, Engineering and Library Science

Session Three

Copyright and Grey Literature
Authorship, Ownership, and Property Rights

Chairperson: Julia Gelfand, UCI
University of California, Irvine

Breakout Sessions:

Track A:

Policy and Management of Grey Literature
Chairpersons: Bruce H. Kiesel, BIOSIS and Masayuki Sato, JST

Track B:

Applications and Uses of Grey Literature
Chairpersons: Carolyn E. Floyd, NASA and Andrew Smith, EAGLE

Product and Service Reviews

Closing Session

Chairperson: Keith G. Jeffery, CLRC
Central Laboratory of the Research Councils

Table of Contents

Acknowledgements

iii

Opening Session

Chair: Dominic J. Farace, GreyNet, Grey Literature Network Service

Julia M. Gelfand - United States 1-11

University of California, Irvine

Grey Literature : Plausible Applications for Distance Education Practices

Helmut M. Artus - Germany 12-24

Social Science Information Centre, IZ

The Internet and the Socio-Structural Change of Informal Scientific Communication

L. Olatokunbo Aina - Botswana 25-31

University of Botswana, UB

Grey Literature and Library and Information Studies (LIS) : A Global Perspective

Plenary Session One

Global Assessment of Grey Literature:

A Brave New World of Topics, Formats, and Uses

Chair: E. Stephen Hunt, NLE, National Library of Education, U.S. Department of Education

Gail McMillan - United States 32-46

Virginia Polytechnic Institute and State University

Perspectives on Electronic Theses and Dissertations

Bonnie A. Osif - United States 47-60

Pennsylvania State University

*International Transportation Literature : An Analysis of Citation Patterns,
Availability and Research Implications to the Transportation Community*

Angela Aceti, Maria Castriotta,

Rosa Di Cesare, and Daniela Luzi - Italy 61-77

ISPESL, Istituto Superiore per la Prevenzione e la Sicurezza del Lavoro;

CNR, Consiglio Nazionale delle Ricerche

*A profile of GL producers in the field of safety and health at workplaces in Italy :
results of a sample survey*

Plenary Session Two

Publishing and Archiving Electronic Grey Literature:
From Production to Full-Text Storage, Retrieval and Distribution

Chair: Eileen Breen, MCB University Press

Michael L. Nelson and Kurt Maly - United States 78-87

NASA Langley Research Center, LaRC ; Old Dominion University, ODU

Preserving the Pyramid of STI Using Buckets

Keith G. Jeffery - United Kingdom 88-108

Central Laboratory of the Research Councils

An Architecture for Grey Literature in a R&D Context

Deborah E. Cutler - United States 109-115

Office of Scientific & Technical Information, OSTI

Grey Literature in Energy : A Shifting Paradigm

Elena Lodi - Italy 116-134

Martin Vesely and Jens Vigen - Switzerland

University of Siena ; European Centre for High Energy Physics, CERN

Link Managers for Grey Literature

Plenary Session Three

Copyright and Grey Literature:
Authorship, Ownership, and Property Rights

Chair: Julia Gelfand, University of California, Irvine

Michael Seadle - United States 135-142

Michigan State University, MSU

Grey Copyrights for Grey Literature : National Assumptions, International Rights

Cees A.T. de Blaaij - The Netherlands 143-155

University of Amsterdam, UvA - Institute for Information Law, IVIR

Intellectual property on the move : Some observations on authors' rights, grey literature, publishers and last but not least access to information in a cyberian age

Dave Davis - United States 156-163

Copyright Clearance Center

The Application of Electronic Copyright Management Systems (ECMS) to Grey Literature

Graham P. Cornish - United Kingdom 164-169

International Federation of Library Associations and Documentation, IFLA

Copyright : Black and White or just making you see Red?

Breakout Session - Track A	Policy and Management of Grey Literature
-----------------------------------	---

Chairpersons:

Bruce H. Kiesel, Publisher of Biological Abstracts and Zoological Records, BIOSIS
 Masayuki Sato, Japan Science and Technology Corporation, JST

Ichiko T. Morita - United States 170-180 41

Japan Documentation Center ; Library of Congress

*Issues in Distribution of Grey Literature : Experience of the Japan
 Documentation Center (JDC)*

Maria do Rosário Guimarães Almeida - Brazil 181-198 102

Univesidada Federal do Maranhão, UFMA

In Absentia: *Control and Access for Grey Literature in Brazil : A Proposal*

Joachim Schopfel - France 199-202 47

Institut de l'Information Scientifique et Technique, INIST-CNRS

INIST : Tracking Grey Literature in a Changing Environment

Atefeh S. McCampbell,

Linda M. Clare, and Scott H. Gitters - United States 203-216 107

Florida Institute of Technology ; Medical System Information Technology Group ;

Lippincott Williams & Wilkins

The Competitive Advantage for the New Millennium : Knowledge Management

Hyun-Yang Cho, Seon-Heui Choi,

Sung-Pil Choi, and Huyng-Sup Shim - Korea 217-226 10

Korea Research and Development Information Center, KORDIC

*Archiving Electronic Grey Literature of Korean Academic Societies
 in Science and Technology*

Breakout Session - Track B	Applications and Uses of Grey Literature
-----------------------------------	---

Chairpersons:

Carolyn E. Floyd - National Aeronautics and Space Administration, NASA

Andrew Smith - European Association for Grey Literature Exploitation, EAGLE

Linda R. Musser - United States 227-229 3p

Pennsylvania State University, PSU

Mine Maps as Grey Literature

Breakout Session - Track B

Applications and Uses of Grey Literature (continued)

Jean-Yves Le Meur and David Dallman - Switzerland.....	230-237	87
European Centre for High Energy Physics, CERN		
<i>Automatic Keywording of High Energy Physics</i>		
Fethy Mili - Canada.....	238-253	169
University of Montreal		
<i>Trends in Publishing Academic Grey Literature : Examples form Economics</i>		
Krystyna Siwek - Poland.....	254-259	69
Information Processing Centre, OPI		
<i>Grey Literature Based Surveys of Information Industry Development in Central and Eastern Europe</i>		
T. Rama Devi and Anil Takalkar - India.....	260-275	169
Centre on Rural Development, National Institute of Rural Development		
<i>Use Pattern of Grey Literature in Rural Development : An Indian case study</i>		
Diane Helmer, Isabelle Savoie and Carolyn Green - Canada.....	276-281	69
British Columbia Office of Health Technology Assessment, BCOHTA		
<i>How do various Fugitive Literature Searching Methods Impact the Comprehensiveness of Literature Uncovered for Systematic Review?</i>		
Hiroyuki Sato, Kenshou Nakagawa and Yoshiyuki Maeda - Japan.....	282-296	159
Japan Science and Technology Corporation, JST		
<i>Directory Database of Research and Development Activities (ReaD)</i>		
Rosa Di Cesare and Giovanni Lazzari - Italy.....	297-306	109
Consiglio Nazionale delle Ricerche, CNR ; Biblioteca della Camera dei Deputati		
<i>Towards integration of information sources on grey literature : A Case Study</i>		
Halima Semra - Algeria.....	307-313	79
University of Constantine		
In Absentia: <i>Internet use : Research of Grey Literature by teachers-researchers at the University of Constantine</i>		

Appendices

List of Participating Organizations.....	315
Author Index.....	319
GreyNet Membership 2000.....	320
Publication Order Form.....	322

GreyNet

Grey Literature Network Service

Your Gateway to Grey Literature

Newsletters

Bibliography

Conferences

Welcome

R & D

Publications

Guide

GL'99

What's New

GreyNet a branch of **MCB** 
University Press

GREY LITERATURE NETWORK SERVICE
Koninginneweg 201, 1075 CR Amsterdam, Netherlands

Tel/Fax: 31-20-671.1818

Email: greynet@inter.nl.net

<http://www.konbib.nl/infolev/greynet>

SIGLE

SYSTEM FOR INFORMATION ON GREY LITERATURE IN EUROPE

**The premier source for the latest advances
in natural and social sciences, technology
and medicine**

- ▶ **PROVIDES:** quick and easy access to European literature, not available through normal bookselling channels
- ▶ **ENSURES:** document/full text supply for each cited item
- ▶ **CONTAINS:** research and technical reports, preprints, dissertations, conference proceedings, etc.
- ▶ **ITS SIZE:** 565,000 records (Dec. 1998)
- ▶ **THE UPDATE FREQUENCY:** monthly

SIGLE is available:

- ▶ **ONLINE:** **Blaise Information Services**
<http://www.bl.uk>
e-mail: blaiseline-helpdesk@bl.uk
Tel.: +44 (0)1937 546430
STN International
<http://www.fiz-karlsruhe.de>
e-mail: hlpdeskk@fiz-karlsruhe.de
Tel.: +49 (0)7247 / 808-555
- ▶ **INTERNET:** **STNEASY**
<http://stneasy.fiz-karlsruhe.de>
- ▶ **CD-ROM:** **SilverPlatter Information Ltd**
Tel.: +44 (0)181 995 8242
Fax: +44 (0)181 995 5159

SIGLE is a product of



an association of leading information and document supply
centres in 15 European countries.

For detailed information and document supply see

<http://www.konbib.nl/sigle>

GREY LITERATURE: PLAUSIBLE APPLICATIONS FOR DISTANCE EDUCATION PRACTICES

Julia M. Gelfand
University of California, Irvine
jgelfand@uci.edu

Paper to be presented at the
Fourth International Conference
On Grey Literature
Washington, DC
October 4, 1999

ABSTRACT: With distance education among the most dynamic elements in post-secondary education today and with changes in emerging instruction technologies breaking frontiers, new interpretations of copyright and creative site licensing of information, grey literature has an equally new potential for use in these structures. The "Innovation Diffusion" common in distance education is a parallel structure to developments in electronic publishing and alternative ways to distribute and treat grey literature. Connectivity remains the major challenge as distance education courses are seeking materials and resources that they can ascertain will be available to their students via course supported software and a host of browsers with few restrictions. Grey literature, due to its wide scope and definition, and oftentimes unrestricted access, lends itself significant use in this arena. There are mutual benefits to the author/editor/compiler, to the faculty, students, the host institution and that library, that may have some lasting impact in how distance education can transition from a dependent extension of traditional classroom based education environments to a widely independent distributor of information. This paper explores different methods of how grey literature can be more widely utilized in such applications and what challenges there are to promote grey literature in what appears to be both asynchronous as well as synchronous environments.

"The time is right for distance learning" begins the 1999 imprint, *McGraw-Hill Handbook of Distance Learning*.¹ What is implied here, is that as we near the millennium, it may be time to reduce the often harsh criticism that is levied against distance education programs and to see why alternative education is so attractive to a growing number of students and campus administrators and leaders. The information age has introduced new methods of delivering instructional materials and content to the student by creating nontraditional classroom experiences. These learning environments have changed rather dramatically as new and emerging technologies can now offer

instruction via a host of interactive channels. These changes create new opportunities and challenges for both student and institution of higher learning.

Distance education is a marketplace in which students are separated from their instructors by time, and/or space. The early history of the distance education movement was born with a dependence on correspondence education, in place since the late nineteenth century, where students received course materials via the mail, acquired materials, completed assignments and returned them to be graded again by postal service. Audiocassettes and broadcast mediums over television and satellite with add-ons such as video-cam recorded tapes, and finally, teleclasses involving the presentation of classes in specially designed classrooms, facilitated the next generation of technologies, followed by the introduction of computers. Classroom architecture was redesigned to accommodate and facilitate the new technology requirements.

When computers came along, in time they came with CD-ROMs, networks, potential for electronic mail, the Internet, mixed media, a vast, rapid emergence of online services using telephone lines and communication software. Thus, what has emerged is a vision of a networked learning environment that "seamlessly integrates voice, video, and data connections between and among learners, instructors, subject matter experts, virtual libraries, Internet resources and support service organizations. The distance learner is at the center of this network, connected with both real-time and non-real-time links to a wide variety of learning resources. The learning resources are themselves interconnected to form a mosaic of mutually enriching elements."²

As more of the general population becomes increasingly computer literate, the windows of opportunity open to invite new ways of doing business in education. E-

commerce suggests different connotations and meanings, but could easily define the new educational enterprises that have developed since barriers were reduced due to a networked environment. Distance education always served the adult learner, the student who was isolated from attending school in a traditional model and had to make alternative arrangements. It may have been because of geography and remoteness, economics, circumstance of needing to work and care for family that forced the issues of creating new models. Regardless, today those reasons are immaterial, because distance education is becoming a choice that extends beyond those considerations, and has proven itself as an alternative to traditional learning styles of being lectured to in increasingly overcrowded halls. Instead, it can offer a more flexible and individualized instruction package to students, with potential for more ways to interact with other students and the teaching faculty.

Computers are the most versatile components of distance education and provide the medium of communication and serve as the stepping stones of interactivity creating asynchronous environments. Software packages allow for institutions to customize online classrooms so students can:

- follow course schedule,
- obtain their lessons,
- ask questions of the teacher,
- share ideas with other students,
- participate in group and team assignments,
- request and transmit texts and graphics from content providers, and
- develop verbal and presentation skills by reading and writing, and engaging in creative measures.

Thus, a virtual classroom now generally offers:

- access to e-mail,
- threaded discussions,
- chat sessions,

- a whiteboard,
- shared applications,
- streaming video/audio and multimedia files,
- course and/or bibliographic management, links,
- interactive ROMs to interact with the course website.³

A recent study reported in the media and conducted at UCLA, says that the stress of staying up-to-date with technology affects more professors than traditional stresses such as teaching loads and publishing demands. The survey shows that 67% of professors are regularly stressed by keeping up with emerging technology, compared with 62% stressed by teaching loads, and 50% stressed by research or publishing pressures. Researchers say fear may be preventing professors from using new technology. Only 35% of professors use the Internet for research purposes, while 38% use technology to prepare presentations for classes. The survey results indicate that colleges should work to improve instructors' computer skills in order to meet the needs of students who have grown up using computers and are comfortable using new technologies.⁴

Many new concepts have been defined as well as the release of products associated with distance learning. They include "thoughtware" which refers to how "knowledge is so important, yet so underestimated and overlooked...that it is critical footing for the installation of new thoughtware in any organization. It's threatening, then disruptive, becomes provocative, then challenging and finally a true agent of change. Learning and development of new knowledge must become a part of every person's accountability."⁵

Some of the difficulties that have been well identified and articulated in scores of articles and books, are that institutions that played significant roles in teaching now must cope and adjust to new operational paradigms. These institutions include publishing and

libraries. Technology has also served these crucial pillars of education well, but challenged them in very complex ways. New relationships emerge with the concept of who is an author at the core of the exchange. Simply stated, the commercial product that has changed most significantly is the textbook. Textbook publishing has moreorless commanded the ultimatum, "publish online or perish."⁶ In addition, servicing of the textbook has also dramatically changed, with electronic versions of library reserve services well in place with easy connections via Library online public catalogs, as well as the role of the campus bookstore, customized textbooks on demand, the course homepage, commercial lecture notes, and an array of other services and products..

The publishing and library communities have been very concerned about the impact of distance education. They have developed many different guidelines to direct the legal, social and ethical distribution of resources to students wherever they are, on campus, off-campus, in distance education or extended campus programs, or in the absence of a campus at all. Servicing course needs in distance education has become one of the fastest growing segments of academic libraries, with scores of newly defined positions for librarians where the emphasis is to provide content and services to the distance learner as seamlessly as to a local student. The Association of College and Research Libraries (ACRL), among the largest divisions within the American Library Association passed a set of revised guidelines for distance learning library services in July 1998.

This document carefully:

- defines all the players in the distance learning context;
- offers a set of philosophical assumptions about distance learning;

- lists how management of such organizations and institutions participating in distance education should consider administering library and information services;
- offers what functions library personnel associated with such a program should be engaged in;
- suggests the kind of facilities that lend themselves to the best blend of library resources and services;
- notes that thorough documentation should support all aspects of the efforts to provide adequate resources, access and information services to support all distance education measures.⁷

It is critical that such guidelines allow for flexibility and changes in the educational setting and promote that local applications should freely explore the best ways to accommodate faculty and student needs for efficient and effective information delivery.

So, where does grey literature enter into this discussion? If we think of grey literature as an adventure that can best be described as:

- 1) it uses knowledge and available data to confront tangible problems
- 2) it immerses the practitioner in real world situations
- 3) it often exists as an ill defined multidisciplinary field lacking clear boundaries,

then the future is clarified and one begins to know the concepts and methods of its construction. Grey literature has been known to enhance the service delivery of information by providing an alternative way of being distributed. It is a tool that builds human capital when used in locating other resources.

Distance education is also an adventure with sometimes a murky future; it has a risky undertaking, is a business venture with a remarkable and exciting experience to offer all who choose to participate. Can't the same be said for grey literature – the prospects are risky, and the journey is a remarkable and exciting treasure trove.

Distance education appears to flourish because students in a variety of settings can now have access to resources for which they have to go nowhere to get. The viability of the Internet as a place where resources are located makes the possibility of offering a wider variety of courses spanning the universe of information greater. With institutions entering into contracts that creatively negotiate the opportunities that libraries can promise for their affiliated students to gain access to materials they need to support their instructional needs, half the battle is won. The other half has to do with ascertaining that copyright compliance is practiced and this can be the more difficult part to achieve and oversee.

Today, grey literature is defined as information output that is not controlled by commercial publishing interests and where publishing is not the primary business activity of the authoring unit or organization. Thus, it is my contention, that grey literature has a natural home supporting distance learning initiatives. Since the span of distance education is now so encompassing and broad, there is even a greater sense of the kind of material that is appropriate to support the learning environment. Students will be conducting more independent research, spending time online and mixing media. As long as copyright becomes the focus and permissions are secured when needed and clearly stated, grey literature may well hold the tenets of intellectual property to serve the range of distance education. Higher education is expected to nourish the creativity, resourcefulness and independent spirit required to solve difficult social and scientific problems. It places an emphasis on independent learning, reasoning and growth.

The sources and examples of grey literature will never be sufficient to address all the demands of curriculum nor be discovered by those who need to use them, however

alternative electronic publishing readily and intensely fuel these needs. If grey literature continues to be well represented in the public domain, it increases the likelihood that it will be found, identified and used more seriously, thus will be cited and referenced in mainstream publications as well as in new forms of grey literature. This will include primary and secondary publishing output. The latter is likely to be more compatible with instructional materials. Examples of this may include interpretations of data sets and statistics, databases of images, concordances of notes and reference works, indexes to important but low-use or low impact works, and student produced work that has not experienced peer review. The latter is a bit scary, because it lacks authority control and contributes to the already disproportionate amount of material on the web with independent domain addresses.

Faculty seeking appropriate and relevant materials may steer students to such grey literature content if it does not require the potential expense, and planning ahead necessary to obtain the proper site license. Experience to date suggests that specialized course homepage has a strong future.

The Report on Copyright and Digital Distance Education issued by the U.S. Copyright Office in May 1999, interprets how the Digital Millennium Copyright Act of 1998 (DMCA) is served by distance education. There are several parts to this report. For the sake of this paper, the most relevant parts are:

- Part I provides an overview and illustrates how current distance education practices operate.
- Part II is the introduction of current licensing practices in digital distance education and with a description of past and current problems and suggests trends.
- Part III describes the status of technologies relating to the delivery and protection of distance education materials.
- Part IV is an analysis of current copyright law and digital distance education.

- Part VI is the examination of whether the law should be revised and what the Copyright Office recommends at this juncture in time.⁸

The Report, hailed as welcoming by the higher education legal communities, responds very well to how the capabilities of digital technologies lend to expanding pedagogical possibilities within the classroom experience. One constant in teaching is to develop critical thinking skills. Encouraging students to seek and retrieve a variety of relevant materials that support their information query can do this. Usually consulting pathfinders, indexing tools or manipulating a variety of metasites and search engines does this. Students then must explore, understand, use, apply and evaluate these resources, and for some people, they may be logical and for others, irrelevant and ambiguous. Alistair Fraser writes, "In the past we relied on words, diagrams, equations and gesticulation to build those models (of how the world works) piece by piece in the minds of the students. We now have a new tool – not one that replaces the older ones, but one that greatly extends them: interactive computer visualization. Today, a teacher can build a pedagogical model, and both student and teacher can interact with it to explore the behavior of the system in a way inconceivable in earlier times."⁹

The increasingly stressful and problematic concerns in the distance education marketplace will not vanish; instead they may become a more complex set of issues, as choice grows and distance learning becomes even more widespread. With students taking degrees from institutions on foreign continents, the institutions of the Ivy League dipping their toes into this once forlorn land, mergers and affiliations with the British Open University, Virtual Universities springing up in every consortia and governed body, with the commercial and private sector well positioned to sell, serve and participate, the blurring of profit and non-profit entities, higher education establishing partnerships with

corporations and industry, it is not clear that we can even rightfully imagine what the full spectrum of distance education will be in three, five and ten years out.

Students will be granted degrees from institutions to which they have only visited online, and our netiquette should improve as we develop casual and long term relationships with those we only know as an e-mail handle. Grey literature will be there, coming up for air, taking a big breath and reinventing itself in a more civilized and communicative foray in as much of a technology neutral environment as possible. Grey literature will experience more company in its domain, with authors internal and external to the academic community wanting to contribute. The dilemmas of how to provide adequate and current enough bibliographic control will keep library staff actively concerned about how to compete with demands, and respond to legitimate and growing fears of unauthorized downstreaming of grey and not grey materials.

I can not help but be inspired by a statement in the conclusion of Albert Borgmann's book, *Holding on to Reality: The Nature of Information at the Turn of the Millennium*, when he writes, "Stories are the spaces wherein pictures and mementos come to life and coalesce into a coherent picture of the past and a hopeful vision of the future. Records in turn keep stories straight and lend them detail. Thus the culture of the word can card, spin, and knit the mass of technological information into a tapestry that is commensurate with reality."¹⁰ My view suggests that is a good way as to how grey literature can be summarized. Efforts to create new learning communities, and to sustain and develop the better elements of contemporary distance education have a strong future as long as the technology can be supported for both asynchronous and synchronous learning environments.

¹ Alan G. Chute, Melody Thompson & Burton Hancock, *The McGraw-Hill Handbook of Distance Learning*. New York: McGraw-Hill, 1999, 1.

² Ibid, 207.

³ *Report on Copyright and Digital Distance Education*. Washington, DC: U.S. Copyright Office, May 1999. Available at <http://lcweb.loc.gov/copyright/> 54-56.

⁴ Angetta McQueen writing for the AP August 30, 1999 about a research study conducted by Linda Sax at the University of California, Los Angeles. No further citation available.

⁵ J. Philip Kirby, *Thoughtware*. Portland, OR: Productivity Press, 1997, 130-131.

⁶ Jeffrey R. Young, "A Pilot Project Adds Security to Electronic Textbook Publishing," *Chronicle of Higher Education*, August 6, 1999, A31.

⁷ ACRL Guidelines for distance learning library services, final version, approved July 1998, available at <http://www.ala.org/acrl/guides/distlrng.htm>

⁸ US Copyright Office, *Report on Copyright and Digital Distance Education*, Executive Summary, May, 1999, I-xxvi.

⁹ Alistair B. Fraser, "Colleges Should Tap the Pedagogical Potential of the World-Wide Web," *Chronicle of Higher Education*, August 6, 1999, B8.

¹⁰ Albert Borgmann, *Holding onto Reality: The Nature of Information at the Turn of the Millennium*. Chicago, IL: The University of Chicago Press, 1999, 231.

SMAISMIRMILMEPOETALEUMIBUNENUGTTAUIRAS

The internet & the socio-structural change of informal scientific communication

Helmut M. Artus

Social Science Information Centre (IZ)

Introduction

To begin with: The *sub-title* of my paper is „*The internet & the socio-structural change of informal scientific communication*“. Please do not urge me to read its *main title* which is not only even more complicated than the sub-title but is definitely unpronounceable and definitely not understandable. Nevertheless, there *is* sense and information in it, even information about an important empirical observation. I shall give the solution and explication in part 2.

My paper deals with the impact of the internet on scientific communication and publication and tries to find out which changes can be expected. At first sight, this seems rather clear and easy, since technical development seems to determine the future.

- So, *part 1* of my paper will deal with the *chances and possibilities opened up by technical development*.

On the other hand, we cannot expect that everything that is *possible* will also be put into *practice*. Not any *vision* is becoming *reality*, neither in politics nor in science. Between possibility and reality, there are things like *acceptance* or refusal which turn out to be really *social processes*, which means that they are *complicated, nonsensical and unpredictable* - at least as long as we did not understand the rules of the game.

- So, *part 2* of my paper has to do with *science as a social system, communication as a social process and scientists as social beings* - and as individual subjects persueing their own interests at least as much as the interests of science.

This part tries to learn from the history of science as well as from sociological theories explaining it. As we shall see, this sociological view can considerably change our expectations...

- ... as will prove *part 3* which tries to summarize the facts and arguments and make a prognosis of what will happen in the area of scientific communication. There, we shall have to discuss the *transition from virtual changes to real changes*.

1. Virtual changes

The internet has been changing our communicative customs in a rather dramatic way, and the end is not in sight by far, yet. Beyond, we can be sure that even the sector of publication - and above all that of grey literature - will undergo severe changes.

Just in order to illustrate this:

- My last five publications which came out this year, were presented in a printed and an internet version at the same time, three of them were additionally passed to about 2,000 addressees by an e-mailing list. The print version, by the way, is sent to 'only' 1,400 people.
- Since about two years, I have been receiving hundreds of e-mails per month coming in from individuals, but in most cases via mailing lists or discussion groups. Since that time, the number of mails I am writing my own has increased as well and is much above the number of letters I used to write.

Such developments have changed our behaviour in an interesting way. Since communication is much easier, we communicate *more*, actively (i. e. *writing*) as well as passively (i. e. *receiving* and - sometimes - even *reading*). Using e-mail or the internet, it has become possible to get information instantly or to receive an answer to your question within hours or even minutes. So why write a letter, look for an envelope, find out the zip code and postage for a remote country and get the stamps you need, and at last drop it into the mail box, if you can make things short and easy: just write a text and send it (usually without reading it again) to the other end of the world just by a mouseclick?!

So I think it is obvious that certain trends are inevitable:

- *Number of (electronic) communication:* There will be much more communication since e-mail and internet make things much easier, much faster and much cheaper.
- *'Density' and intensity of (electronic) communication:* The number of people participating in communicative or discursive connexions as well as the 'density' of communication will grow, i.e. the number of contributors as well as the number of contributions. The net of questions and answers and statements will get closer and more intense. The reasons - besides the ones already mentioned - are that the anonymity of such connexions diminishes inhibitions and encourages spontaneity, so that a lot of people do participate at all or more often.
- *Acceleration of (electronic) communication:* Communication will become much faster which does not only mean that it is technically transmitted in a much shorter time, but as a result, discourses will be transacted in a shorter time. Either you participate in a topical discussion within a few days - or you are late and the topics discussed have completely changed meanwhile. Ironically just the immense saving of time causes an enormous time pressure. So it could be asked whether - as far as communication is concerned - the internet really is a time saver or maybe rather a time consumer!? In fact, it saves time - but we are not free to make an optional usage of it because by that same time saving process it imposes time pressure. I think, most people talking about time saving are not aware of this dark side of the net...

Similar trends will occur in publishing:

- *Increase of (electronic) publications:* We can be sure that the number of (electronic) publications of any kind will dramatically increase. At least in the industrialized countries everyone has his own computer, mostly with an internet access. So anybody can create his own homepage and bring his non-commercial one man publishing house into the world wide web so that the whole world has a chance to download and read his papers.¹
- *Acceleration of electronic publication:* We can count on an acceleration of (a part of) publications as well, since the internet allows the avoidance of the time consuming (and - as seen from the author's point of view - risky) referee system deciding about the acceptance or refusal of his paper. And at the same time, the gap between the technical processes of producing and publishing (printing) is removed since there is no considerable expense to prepare a text file for 'publication' in the net.

¹ A few days ago I read about the representative of a famous German publishing house, saying that the 'end of the Gutenberg galaxy is not yet in sight': Publishers still earn their money with printed

- *Extension of our understanding of literature:* Short papers for which there was no chance of publication in the past could get a chance now to be published in the internet. Thus, a new genre could be created, short and very grey treatises which used to be kept under lock and key in former days or were only communicated in small circles with restricted access. So the size of a paper ceases to be a criterion for publication. It is no longer the number of pages that decides whether a paper is (or can be) 'literature' or 'nothing'.
- At the same time (and for the same reasons as mentioned with communication), the inhibition of publishing papers will be reduced. This will lead to an increasing number of people publishing at all as well as of the average number of their publications.

If we add up one and one, we must draw the conclusion that boundaries are shifting, be it those between communication and publication which used to be clearly separated facts, be it between grey and white literature, be it between grey literature and what I would like to call *dirty-grey literature* (*dirty grey*, not *dirty literature*!), by which I mean: communication which is written and not private, but is still beneath the level of grey literature, e.g. has no institutional background, no formal 'bibliographic' (or *internet-ic*?) properties etc. So the whole sector of scientific communication - be it private, informal or formal - is in motion, a proper revolution seems to be in full swing. It seems, that there will be a competition between (printed) *grey literature* (GL) and *electronic literature* (EL). If we only gaze at the technical facilities and possibilities, the fight is already decided: *EL beats GL by technical knockout!* But don't abandon hope! The next chapter deals with social facts, and that can change a lot! So be of good cheer!

As a consequence we could expect that the greatest loser of the *internet revolution* is exactly the kind of literature we love most, i.e. grey literature. So perhaps we better should write an epitaph on a species of literature becoming extinct.

Functions that used to be fulfilled by grey literature - especially the quick and uncontrolled, informal dissemination of facts and figures, of unfinished ideas and unaccomplished interpretations - can be fulfilled by the internet as well and in many cases even better, much better.

Nevertheless, I do not believe in the sudden death of grey literature but quite in contrary count on a glorious resurrection of that kind of literature, although in electronic shape. I believe that this *come back* will be really enforced by the processes happening in the internet. We shall see - in parts 2 and 3 of this paper.

2. A sociological view on science & scientific communication

Let me now come to the solution of the riddle given in the title of my paper.

a) A historical episode

As I admitted before, the title is definitely *unpronounceable* and not understandable. Nevertheless, there *is* sense and information in it, even information about an important empirical observation.

The tape-worm-like word is a so-called *anagram*, which means that the author - a certain *Galileo Galilei* - has taken a whole sentence indicating a brandnew observation of his and arranged the letters in a completely different order thus definitely excluding any understandability.

The original sentence - in Latin, of course - reads

„Altissimum planetam tergeminum observavi“

which means:

“I observed the highest planet (=Saturn) in threefold form.”

I think, even this text needs another translation to be understandable. *Galileo* wanted to express that he had discovered *two moons* of Saturn. Actually, he had observed the *ring system* surrounding the planet. *Kepler*, the contemporary genius and (at least in *Galileo's* eyes) rival tried weeks and weeks to find out what *Galileo* could have expressed by his anagram, but failed.²

What happened here?

To understand this historical episode, which took place in the year 1610, we have to keep in mind that the reward system of science is somewhat different from that in business or in other societal sub-systems. There is neither *money* as a reward nor something like *intellectual property*, not even for the one who made a discovery. There is only *prestige and esteem* - but only for the one who is first! Only priority counts. There are no silver medals in the race of scientific priority. Either you win gold or you lose.

In such a system the indispensable interest of a scientist must be to be the first - and to keep a safe lead over all competitors. Actually, there is a remarkable contradiction (or at least something very close to a contradiction) concluding from the scientific norms and rules: If a scientist makes a discovery, he is first, but he can only earn the grapes of his

² Actually, he found a solution which he himself scornfully called a 'semi-barbaric Latin verse' and which, besides, was wrong: „*Salve umbistineum Martia proles.*“ - „*Salute, glowing twin, sprout of Mars.*“ He thought *Galileo* had discovered a *moon of Mars*.

performance if he publishes his findings. Doing so, however, he gives his rivals the chance to surpass him. So by publishing his discovery he imposes a new stress on himself and can no longer be sure to keep the lead

This was *Galileo's* situation, but he found a very interesting way of securing his priority without, however, allowing his rivals - above all *John Kepler* - to derive any profit from it. His idea was to "publish" the completely un-understandable anagram.³ By decoding his message, he could at any time prove his priority, and by not explaining its meaning he could be safe from *Kepler* and others.

b) The "storybook version" of science

This is much more than just a historical episode but leads us straight into the social life of science, into the very human behaviour of scientists as human beings, and into the problematic and sometimes contradictory connexion between science as a *cognitive* and science as a *social* system. As *Galileo's* behaviour proves, there can be reasons for retaining information and for not or not really publishing them. The famous episode is a good example for certain contradictions between personal (subjective) interests of scientists as individuals and social norms of science as an intellectual connexion if not even for contradictions within the system of scientific norms itself.

I shall try to explain this in a few words.

In general, science is depicted as a *cognitive connexion*, i.e. an arrangement aiming at knowledge only. Each and everything happening there has a scientific, i.e. a cognitive, intellectual or theoretical meaning, individuals or groups acting within this system are subject to the idea of creating, testing or communicating scientific knowledge, of producing progress and so on. Although human beings, scientists do nothing than execute the rules of appropriate and methodologically correct research. As long as they act within the social system of science they guide their whole action along scientific knowledge. Actually, there are deviances from the rules of appropriate professional conduct, but these are just individual deviations which are sanctioned by the scientific community or even by juridical courts.

³ Of course, *Galileo* did not formally *publish* the anagram expressing his discovery, but gave it to the Toscanian ambassador at Prague (*Kepler* was astronomer at the Emperor's court at Prague at that time). Three long month later *Galileo* explained the anagram's meaning - but not to *Kepler* or any other scientist, but to Emperor *Rudolph*; p. 382 in: *Arthur Koestler: Die Schlafwandler. Die Entstehungsgeschichte unserer Welterkenntnis* (first published in English: *The Sleepwalkers*, 1959), Suhrkamp: Frankfurt/M. 1980

What I have described here is what *Blaise Cronin* has called the '*storybook version of science*'. Scientific fairy tales. But nevertheless, this is the image and self-image of science, and even the sociology of science who should know better (and look at its object much more accurately, with skepticism and sense for realism), - even this sociology of science often fails to look at science as *it is* and instead describes science as *it should be*.

c) A realist version of science

From a sociological (or simply *realist*) point of view this is sheer nonsense. From its very beginning, science has been a *social* connexion realised by human individuals, acting together (or against each other) in social groups, under social aims, values and norms. If we look at scientists, we see real social subjects, i.e.: people who earn their living by doing their jobs like bakers or engineers or shoemakers. So job and money are inseparable parts of science as a (really existing) social system and of a scientist's life. And scientists - as really existing people, not as bloodless abstractions as presented in the *storybook version* - , scientists are real human beings with individual interests and preferences and convictions and goals etc. of any kind - scientific as well as political, religious, philosophical, sexual, social and so on. As real human beings, they also have their characters and feelings or, in other words, a *psyche*. They are as noble, decent and generous or mean, evil and envious as any other people, they are jealous or cordial, they are intelligent or - sometimes - stupid, they are healthy or mentally ill as any other people. And last but not least they are egoistic, selfish, and vain. They do not only want money and awards, prestige and esteem, but also admiration, not to talk about sex. (But this is another story, that you cannot find in written descriptions but only in the *oral history* of science .)

If we accept this - and I think there is not one reason for denying it -, if we accept this we cannot exclude really human features from science but have to admit that scientists - as any other human beings - act in a way that is a mixture between professional rules and norms on one side and their individual characters and interests and psychological predispositions on the other. If you do not accept this, please remember the *Galileo* episode giving a very good example for this mixture. It does not only prove that a scientific genius as *Galileo* can have a rather bad character (which, to my knowledge, he had), but that individual features and interests can interfere with the scientific process so that science cannot be conceptualised without regard to the scientist as a real human.

d) Some remarks about social inequality in science

One should assume that scientists would appreciate any innovation promising to improve scientific communication and the dissemination of information and knowledge, such as the internet. But as we have seen, there are reasons for not communicating, for not revealing information, for placing personal interests above the interests of science and knowledge. These reasons can be cognitive or social, they can be compatible with appropriate scientific conduct and they can be very personal, selfish and contradict any scientific norm and value.⁴

The role of the scientist as an individual and as a really social being is one side of the problem. The other and much more important side is that modern science in the course of several centuries has been developing a comprehensive set of goals, values, norms and sanctions as well as social structures and of functions that have to be fulfilled, if science should work appropriately. Thus, science as a *cognitive connexion* is based on science as a *social connexion*. The social system deserves its goals and values to the *cognitive* definition of science, but on the other side, the cognitive connexion can be realised if and only if the social system works - and only in as far as this social basis works.

At first sight, it might be self-evident, that science is a deeply democratic domain, where equals interact in an ideal way. Actually, however, *the idea of science requires differences*: There must be something like social stratification, of inclusion & exclusion, of *social inequality*. If *prestige* and *acknowledgement* are the valuta of science and its reward system, as I have explained above, then all factors providing such necessary social differences are necessary for the functioning of science:

- There are, for instance, *competition and rivalry* for priority and acknowledgement; these are social processes, but their indispensable function is a cognitive one.
- Not publishing itself grants prestige, but publishing in a journal or a publishing house with high reputation. And the most reputed journal or publishing house is the one with the highest rate of refusals or with the highest barrier to access. The greater the difficulties, the higher the prestige for the happy few who succeed and whose articles are accepted.

⁴ During the last decades, commercial reasons have increased. *Science based industry*, I presume, cannot follow exactly the same rules as basic research, and the same holds for research obliged to the sector of defense. As we know, considerable parts of e.g. informatics have not been publishing in the 'official' (open) publication system of science for years and years, but in a parallel one hermetically closed against the rest of the world. Here we find another version of *Galileo's public-secret-dilemma*: Scientists doing top secret research work can *publish* (!) their results - but only within a publication system which as a whole is a secret itself. So *keeping a secret and publishing it at the same time is possible*. Of course, this is a perversion of science.

This is not because scientists are secret masochists but because these are fundamental differences that make science work.

It is self-evident that those who have undergone all the problems, troubles and toils of the long and intricate way to the top, do not have the slightest interest in changing the system to whom they owe their prestige. This may be considered to be a more *psychological* and subjective attitude, but it isn't. The mechanisms providing the social inequality of science are basic. Without them, science would give up the idea of elites (and the difference between the elite and the scientific crowd, between the genius and the simple and unimportant researcher etc.). There is nothing else that could provide a substitute. Then, science would turn out to be a social system like any other without anything special and, above all, without any chance to claim a superiority of its special kind of knowledge as compared to any other.

So we can start from the assumption that these functions which are of fundamental importance are subject to a *meta-stabilization* warranting that their shape may change, that their outer appearance may differ from time to time, but that *if* they are replaced, they are replaced by other 'phenomena' fulfilling exactly the same functions. Anything else would mean giving up fundamental principles of science.

So despite any possibilities offered by the internet, we can expect that this "functional nucleus" of science, warranting social inequality of a certain kind, will not be touched, diminished or even destroyed or replaced by something else. So we can expect that the *appearance* of scientific communication and publication in the age of the internet could change, but at the same time we can expect that on the *functional* level (where all important things happen) *only little will change - if anything at all*.

3. Real changes: What do we have to expect?

Now that the sociological questions are somewhat clear, we can try to draw the conclusions. So, let me now come to the third and last part of my paper and to the question: *What do we have to expect?*

Let us at first ask what is going to stay although it may change its outer appearance:

There is no doubt that a lot of communication, grey and even white literature will find its way into the internet in the near future. I would not argue about that. But my *thesis* is that despite any changes caused by this new technical medium, there will be only little really *substantial* change. Scientists will *use* the internet but will not *subjugate* to it. They will by no means change their customs, their patterns of behaviour or even the social structures of

science or scientific communication. So the changes to be expected, I think, will be much less dramatic than they seem (or threaten) to be.

After all I have said, I think the following points are clear:

- The fundamental *difference between uncontrolled (public) and controlled communication* which is restricted to small groups and personally known participants will remain. It is an illusion to expect that the leading scientists and experts would open their communication channels to anybody.
- The same holds for the *difference between controlled and uncontrolled access to publication media*. Of course, anybody can create his own homepage and offer his private publishing house; in so far the internet will change a lot and provide more democracy. But not only *prestige* but also other scientists' *interest* (!) will now as ever be related to media with strictly controlled access. So the mere mass of additional electronic publications will neither question the boundary between restricted and unrestricted publication, nor will it change the course of the publication process. The transition from print to electronic publishing will not change anything of such a fundamental importance as the *social stratification of science*, i.e. the difference between top scientists and unimportant project researchers (and all qualitative levels in between).
- Also, the difference between *preliminary and final publications* will not be questioned, although the *channels* or media of information could change. As long as prestige has to do with priority, it is necessary that scientists can publish their findings before they are able to give their publication the final shape and present it in perfection. I am sure that they will find new and surprising ways of using the internet for this very purpose. But I am sure as well that the basic difference between such pre-information just in order to save a priority and the formal publication will remain the same, even if - at the surface - things look different. What was important for *Galileo* will be important for his colleagues in the internet age as well!

At first sight, it may look seductive to publish without any restriction, without a peer review system, without the risk of being refused by the editor. But this argument would hold for grey literature as well. Why did grey literature not replace formal publication? For the same reasons, we could ask now: Why will the unrestricted internet site not replace other publication forms which may be electronic but will be strictly controlled in any case? The reasons are the same in both cases: Neither grey literature nor the free internet site can create social inequality, which - as we have seen - is an indispensable function within the social system of science.

So let us ask the second question: *What will change?*

- Without any doubt, a lot of communication and publication is going to be electronic. The advantage of the internet is too obvious to waive its possibilities. Some of them, I have already described above: Communication will grow in number and become closer and more intense.
- A certain type of information, communication and publication will go over from print to the internet, notably *information with trivial contents*. There is a lot of information which is necessary for scientists although it is not scientific at all. Take for instance information about institutes or research projects, about the controlling gremia of research institutes, listings of approaching international conferences, brochures with new publications and so on. In the past, most of such information was published in form of prospects, in the future, they will be disseminated via internet pages. Trivial *communication* will use e-mail (as it already does).
- One of the most interesting and important things is not to stay but has still to *develop*. The internet is not only a medium for arbitrary usage, but a kind of *complex service center*. In a fascinating way it could link the different parts of the information sector which belong together but for historical reasons have undergone separate developments. These are commercial publishers, research institutes producing grey literature, libraries keeping literature in stock and lending it, book sellers and - last but not least - database producers and information brokers. Using the internet it could be (or better: *become*) possible to make an online database search, then - within the same online session - order a copy of an article from a library and download (for money) another one from the commercial publisher's server, order a book from a book seller and send an e-mail to the research team that has published some grey material of interest. This, I think, is what many scientists have been dreaming of for generations.

But what about *grey literature*?

- A lot of grey literature as well as of formal publications will be published electronically. But despite the transition to the internet, the *difference between grey and white literature will be maintained*. The financial interests of commercial publishers are as important as the scientists' interest in the *difference between refereed and non-refereed publication*. Of course, many new publications of any kind and quality will be accessible via internet - but be sure that a cash-box will be installed if the information is really worth reading or downloading. Trivial, unimportant, uninteresting and unrefereed literature will perhaps be free, but *anything of quality and interest will have its price*. The mode of account will of

course be electronic. So the invoice may be only *virtual*, but the money you have to pay is damned *real*!

- As pointed out above, the boundaries between communication and publication will be contested. Notably at the point where (at the time being) grey literature has a more *communicative* intention and aims at *feedback* more than at informing colleagues, - notably at this point, the demarcation between communication and publication will become blurred, texts will be brought into a *discursive* shape rather than being given formal preliminaries etc., so that they can be correctly cited. Such a development, by the way, should be welcomed since communication is one of grey literature's major functions. And if we keep in mind, that science as a whole is a discursive connection, we should be glad that the (to a certain degree) artificial boundaries between communication and publication are pulled down at least at one point. (Nevertheless, even the contributors of such communicative non-papers will take care that their personal performance can be identified and honoured. I think it was the famous sociologist of science *Robert K. Merton* who said that mere *participation* in the game of science was enough appreciation, so the authors should remain *anonymous*. But of course he knew very well that this was just a fiction since scientists are real human individuals with real human interests and vanity.)
- As mentioned before, grey literature will not be replaced by anything else. But on the other side, the producers of grey literature have to expect the most severe problems. More than any other *genre*, grey literature is threatened by an immense number of completely unrefereed, uncontrolled and undiscussed papers from '*one-man-institutes*' and '*one-man-publishing-houses*' - consisting of 1 person, 1 computer, 1 internet interface and at least 1 paper to be published. So if grey literature does not want to be confused with an uncountable number of (more or less) worthless or at least dubious papers, thus losing its reputation as a serious scientific information medium, then it has to draw consequences.

Ironically, the future of grey literature is not more freedom and less control, but quite in contrary a sort of *new institutionalization*. Grey literature, notably *electronic* grey literature - which, of course, is disseminated via internet -, *must* care for a clear and unequivocal demarcation against a cosmos of anything that is beneath its level. This, I think, is the *price for survival*. Grey literature (in our sense) must not be refereed (thus losing its identity) but has to get something like a *brand or a trademark* proving that its origin is a respectable institution or project and making clear that we do not have to do with "*dirty grey literature*" or "*private homepage literature*" with free access to each & every one.

Conclusion: *G. L. is here to stay*

The internet has been changing all our communication and publication habits, that is true. But it is true, either, that also in the internet *man* is measure of everything.

Lots of literature will be distributed via internet. We can expect some substantial changes, but only in a few cases. In most cases, however, the internet will only be used as a new technical medium for selling old wine in new bottles, without changing the substance of its contents, the functions or the publication and (partly) communication system of science.

And so, after all, I think that there is no reason to worry about grey literature's future.

GREY LITERATURE AND LIBRARY AND INFORMATION STUDIES (LIS): A GLOBAL PERSPECTIVE

Prof. L. O. Aina

Department of Library and Information Studies

University of Botswana

P/Bag 0022

Gaborone, Botswana

Tel: 367-359490

Fax: 267-585098

Email: ainalo@noka.ub.bw

Abstract

A lot of studies on grey literature abound in the literature. Most of the studies, however, relate to disciplines other than LIS. The main objectives of this study are to determine the focus and trends of grey literature publications. Other objectives include determining the proportion of grey literature published that focuses on national, regional and global issues, identifying the various formats of grey literature prevalent in LIS, and the major sources used by information professionals in publishing their grey literature research. In order to carry out the study, an analysis of grey literature publications as recorded in Library and Information Science Abstracts (LISA) from 1993 to 1997 was carried out. During the period considered, using the free text option, 179 publications had anything to do with grey literature but only 101 or 56.4% were focussed on grey literature, thus only 101 abstracts were used as data sets for this study. The findings revealed that the major areas of concentration during the period considered were collection (46.4 %), information technology (18.2%), bibliographic control (17.3%) and characteristics of grey literature (13.9%). A large proportion of the researchers focussed on issues affecting Europe (39.6%). This is followed by global issues which constituted (33.7%). The majority of the research findings were published as conference proceeding (63.1%). Journal, as a medium of publication constituted (30.1%). Other media used were reports and books. International Conference on Grey Literature Proceedings constituted the major source for reporting grey literature as it accounted for (57.3%). The study recommends that the International Journal of Grey Literature recently launched should be given maximum support so as to serve a major medium for publishing research findings in grey literature. Other recommendations include the need for more research in grey literature in the other continents especially Asia where less than 5% of the grey literature published emanated from Asia, the need to research into other aspects of grey literature besides collection, characteristics and bibliographic control, as these three areas constituted more than 80% of the papers published.

Introduction

At the third International Conference on Grey literature held in Luxembourg, in 1997, a consensus definition of grey literature (GL) was adopted. This comprehensive definition states that "Grey literature is that which is produced at all levels by government, academia, business and industries, both in print and electronic

formats but which is not controlled by commercial publishing interests and where publishing is not the primary activity of the organisation”(Farace, 1997). In spite of the fact that it is not commercially available, it is generally recognised that it is the first port of call, when reporting important findings, be it research, corporate or even individual activities. It is therefore not surprising that GL comes in the form of policy documents, corporate reports, conference proceedings, theses and dissertations, feasibility studies, working papers etc. There is no doubt that grey literature is valuable in research activity, policy and decision making. In some disciplines, it constitutes a sizeable proportion of the literature. According to Chilag (1982) much of knowledge and information will never appear in any other form except grey literature. A study in grey literature use on agriculture revealed that as much as 14% of grey literature was used for research in agriculture in Nigeria (Aina, 1987). The past three international conferences on grey literature have brought out clearly the need to manage grey literature just as the white literature. Given the fact that the paradigm shift is from ownership to access, many research workers have emphasised the need to market, advertise and disseminate grey literature so as to provide greater access, because of its importance in the generation of new knowledge.

Library and information workers constitute one of the important stakeholders in making grey literature accessible to all who those who need it. Many studies have been carried out to show the use and importance of grey literature in many disciplines other than library and information studies (LIS). This present paper aims to find out how much grey literature is used in reporting research and corporate activities in library and information studies. The main objectives of this study are to determine the focus and trends of grey literature publications in library and information studies. Other objectives include determining the relative proportion of grey literature published, that focuses on national, regional and global issues; identifying the various formats of grey literature prevalent in LIS, and the major sources used by information professionals in publishing their grey literature research.

There are a variety of ways for studying the use patterns of grey literature by professionals in any discipline. One way of carrying out the study is to survey researchers, using questionnaire and interview as survey instruments for collecting the data. This will involve seeking their opinions on the frequency of use of grey literature as an important medium for reporting research findings.

Another alternative is to compile a comprehensive bibliography of the writings of professionals in any discipline and then analyse it bibliometrically to find out the proportion of what constitutes grey literature vis-a-vis white literature. A more direct way, however, is to use an existing data base on library and information studies

research, which can then be analysed. This is the method adopted for this study.

In carrying out this study Library and Information Science Abstract (LISA) database was used as the source data. This is based on the assumption that no abstracting agency will abstract a worthless document given the time and efforts needed in writing an abstract for a document. Thus, any important document published on grey literature will be covered by LISA.

LISA is the principal abstracting agency for abstracting documents in the information profession throughout the world, and it was the natural choice for this study. The assumption of this study is that LISA accurately represents the research strength of every country of the world in the information profession and whatever is covered during the period considered accurately reflects the output of each nation. The electronic version (LISA Plus) 1998 was used in identifying all the literature pertinent to all the various countries and regions of the world. Using the search facilities of LISA Plus, the free text and country of publication fields were searched for all countries in the world. The countries were then grouped into regions. For each paper covered by LISA, the subject of research, country/region to which the research was addressed, the format of publication, and the name of the journal, if the format of the paper was a journal, were recorded on a cataloguing card. In order to carry out the study, a survey of grey literature publications as recorded in Library and Information Science Abstracts (LISA) from 1993 to 1997 was used.

Results and Discussion

An analysis of LISA during the period covered, revealed that there were 179 publications that mentioned grey literature, although only 101 publications were concentrated on grey literature. Thus, 101 publications were used as the database for this study.

Table 1 provides the geographical coverage of grey literature for the different regions of the world. The various countries were grouped into four regions - namely Africa, Asia, Europe and America were identified. Findings reported under global reflect the documents that address issues that go beyond national and regional boundaries. From the table, it is obvious that Europe (39.6%) is the single largest region that has made more contribution on GL than any other region. This is not surprising given the fact that research on grey literature use has been extensive in this region. European Association for Grey Literature Exploitation (EAGLE) has been the principal agency for promoting the use of grey literature in the region. A sizeable proportion of the papers focussed on global issues (33.7%). It is also significant that more papers were recorded in 1994 and 1996 which tally with years

the GL proceedings were published.

Table 1 : Geographical Coverage of Grey Literature Publications (1993-1997)

Distribution	1993	1994	1995	1996	1997	TOTAL	%
Europe 39.6	5	16	7	10	2	40	
Global 33.7	1	12	3	8	10	34	
Africa 13.9	1	6	1	6	-	14	
America 7.9	-	3	1	4	-	8	
Asia 4.9	-	3	-	2	-	5	
TOTAL 100.0	7	40	12	30	12	101	

Table 2 Formats of Grey literature Publications (N=103)

Format	Number	%
Proceedings	65	63.1
Journal	31	30.1
Report	4	3.9
Book	3	2.9
TOTAL	103	100.0

Table 2 reveals that there were some publications that appeared in more than one format. Some papers which have earlier appeared as proceedings and which were

later published as journal articles, hence the number of publications is two more than 101. However, from the table it is obvious that most of the GL publications in library and information studies appeared as proceedings (63.1%). A further analysis of the specific sources used for reporting GL as revealed in table 3 shows that 75.7% of research findings in GL were reported only in seven specific sources and that the *Proceedings of the International Conference on Grey Literature* published in 1994 and 1996 by the Grey Literature (Service Network) the Netherlands, which is the major sponsor of the biennial conference on grey literature was the major medium for reporting grey literature. Majority of the publications appeared in the the International Conference on Grey Literature Proceedings (57.3%). Journals (30.1%) as format for reporting grey literature constitute a significant proportion, although the publications seem to appear in a variety of journals. Twenty-one journals were responsible for publishing 31 articles. With the exception of *Publishing Research Quarterly* (5), *Bolletin AIB* (4) *Serial Review* (2), *Documentation et Bibliothèques* (2), *Interlending and Document Supply* (2), only one article each appeared in the remaining 16 journals.

Table 3: Major Specific Sources for grey Literature (N= 103)

Sources	Number	%
Proceedings of the International Conference on Grey Literature	59	57.3
Publishing Research Quarterly	5	4.9
Bolletin AIB	4	3.9
IATUL proceedings	4	3.9
Interlending and Document Supply	2	1.9
Documentation et. Bibliothèques	2	1.9
Serial Review	2	1.9
Total	78	75.7

Table 4 shows the distribution of the subject content of GL publications during the period considered. The total number of publications is 110 as opposed to 101. This is because some publications cover more than one subject area. The table shows that researchers in LIS focussed more on collection of GL more than any other area. This is not surprising given the fact that a major function of the information professionals is to collect GL. Other significant subject areas focussed on include information technology which is expected to be the main plank for making GL accessible. Of course bibliographic control continues to be a major area of interest to information professionals since there is a need to move away from ownership to the provision of access. The emphasis on characteristics of GL is not misplaced because it is perhaps important to identify its features so as to be able to define GL appropriately, hence the consensus definition arrived at the 1997 International Conference on Grey Literature.

Table 4: Distribution of Subject content of Grey Literature (N= 110)

Subject	Number	%
Collection	51	46.4
IT	20	18.2
Bibliographic Control	19	17.3
Characteristics	14	12.7
Co-operation	2	1.8
Communication	2	1.8
Marketing	1	0.9
Copyright	1	0.9
TOTAL	110	100.0

Conclusion and Recommendations

From the study it is apparent that there are few outlets for reporting research

findings in LIS. Unlike in most other fields where journal format constitutes the principal medium for reporting research findings, the opposite is the case with GL where proceedings of conferences provide the medium for reporting GL. This further makes access to GL even within LIS more difficult. The study also confirms the leading role of GreyNet (Grey Literature Network Service), Netherlands for being the principal agency for promoting research in GL. The study also reflects the need for the consensus definition of GL as adopted at the last GL conference. This is because most of the areas concentrated on, during the period considered, were collection and characteristics of GL. However, with the adopted definition, it is expected that the focus on these areas will shift to access. Europe is the most active region for the promotion of GL as already shown in table 1. Surprisingly Asia contributed less than 5% to GL research during the period considered.

The recently launched *International Journal of Grey Literature* has really come at an auspicious time, it is hoped that the journal will be supported by researchers in the field. Given the fact that access rather than ownership, is the norm, now it is expected a lot of research will now focus on how information technology can be utilised to ensure that GL becomes more accessible. It is hoped that other regions will focus more on GL just as Europe.

References.

- Aina, L.O. (1987) Grey literature and Agricultural Research in Nigeria. *IAALD Quarterly Bulletin*, 32 (1)47-50.
- Chilag, J. (1982) Non-Conventional Literature in agriculture-An Overview. *IAALD Quarterly Bulletin*, 27 (1) 2- 7.
- Farace, Dominic (1997) Third International Conference on Grey literature held in Luxembourg, 13-14 November 1997.

Perspectives on Electronic Theses and Dissertations

Gail McMillan

University Libraries, Virginia Polytechnic Institute and State University

INTRODUCTION

This paper provides three perspectives on electronic theses and dissertations (ETDs): the student's, the faculty member's, and the librarian's. These might also be described as immediate, near term, and long term perspectives. While this paper describes largely recent events, it also draws on the ten year history leading to the Networked Digital Library of Theses and Dissertations (NDLTD), especially the effects of the ETD requirement at Virginia Tech since January 1, 1997. A surprise to some, a revelation to others, but it is a fact of life now for the students, faculty, and library at Virginia Polytechnic Institute and State University (Virginia Tech). As a result, we have gathered a body of information, factual and anecdotal, that describes the impact of ETDs on many cultures both within and outside the academy. This paper presents the ETD initiative from these perspectives.

STUDENTS' PERSPECTIVES

When graduate students at Virginia Tech submit their ETDs, they are asked to respond to a nine-question online survey. The following summarizes the surveys received from 455 graduate student authors submitting their ETDs between February and mid-August, 1999.

When they needed help with their questions about ETDs, 60% of our students sought personal assistance, turning to their friends, committee members, or staff at the library's New Media Center (NMC). The most frequently consulted source of personal assistance to authors' questions about ETDs was their friends (42.6%). Graduate students' consulted their committees nearly 8% of the time, and the NMC staff worked with nearly 10% of our ETD authors. It is the only computer lab that has staff trained and assigned to help its users. Other open computer labs on campus have computers and software, but staff monitor the facility and are not assigned to assist users.

When they wanted answers to their questions about ETDs, three-fourths of our students also turned to the Web. 74.9% said they consulted the VT Web and a few (1.1%) went elsewhere on the Internet. We asked authors who used the VT Web (<http://etd.vt.edu>) how helpful they found it, and 72.2% found it helpful or extremely helpful. Only 1.6% said it was extremely unhelpful and 6% said "unhelpful." 13.9%

found the Web neither helpful nor unhelpful. Staff working with the Digital Library and Archives are reorganizing and updating the VT Web so it will continue to be useful.

Another source of helpful information was attendance at the ETD workshops organized by the Graduate School. Presented twice a semester and once during each summer session, the workshops are held at a variety of locations and at times meant to appeal to students. These workshops have two goals: to introduce students (and faculty who may attend) to the ETD process, and to provide help and advice for preparing ETDs. The workshops are presented in two segments: the first portion is one for the "novice" and the second is for those with more advanced skill levels. The former emphasizes some typical word processing steps, such as using page breaks and how to automatically insert page numbers, as well as preparing tables of contents and integrating portrait and landscape page layouts in the same document. The instructor also covers the steps from the word-processed document to PDF (portable document format) and submitting an ETD through the Web to the Graduate School.

The ETD workshops, however, were not as successful as the Web site in providing graduate students with assistance, because only 21% said the workshop was a source of answers to their questions about ETDs. Of those who said they attended the workshops, over half, 56.9%, found them helpful or extremely helpful. A large percentage, 34.1%, was neutral on this issue, while less than ten percent did not find the workshops helpful.

Whether they got help through personal contact, the workshops, the Web, or elsewhere, over half (256) of our graduate student survey respondents were pleasantly surprised that turning their theses and dissertations into portable document format (PDF) files and submitting them electronically was easier than they expected. Over one fourth, 28.7% reported that creating PDFs were somewhat less difficult, and 27.6% found PDF conversion much less difficult than they had expected. Unfortunately, we still had nearly one-fourth of our students, 106 out of 453, who found PDF more difficult than they had anticipated. "Somewhat more difficult" was the response from 15.7%, and 7.7% said it was "much more difficult." Based on discussions with New Media Center staff, the ETD workshop instructor, and the Dean of the Graduate School, it appears that older graduate students have less experience with computer technology and, therefore, encounter more problems converting their

theses and dissertations to ETDs. Others that have difficulty are, not surprisingly, those who wait until the last minute and find they lack the time to learn to use the technology. One-fifth were not surprised one way or another by the conversion to PDF process.

The results from our graduate student authors were similar when surveyed about the ETD submission process. Over half, 56.5%, found it less difficult than they expected, and nearly one-fourth found it more difficult. As with converting files to PDF, one-fifth found it neither more nor less difficult to submit their ETDs than they had expected.

Where were the VT graduate students when they submitted their ETDs and completed these survey questions? Nearly two-thirds were working on campus: 28.1% were working unassisted in a computer lab while 6.6% were in the New Media Center. Almost one-third, 29.8%, submitted from their campus offices. Nearly 20% submitted their ETDs from their residences; 18.4% were in their off-campus residences and a few, six or 1.3%, submitted from their campus dormitory rooms. 15.8% submitted their ETDs from elsewhere.

What kind of computers did these ETD authors use? The vast majority used PCs: 83.7%. Mac users comprised 13.7% of the survey respondents. Nearly a dozen others, 2.4%, used Unix-based operating systems.

The final questions of the nine asked of our graduate students submitting their ETDs, are very telling ones. In the last two questions the graduate students report their plans to publish from their theses and dissertations, and if they restricted access to their works, who advised them to do so.

Virginia Tech's ETD authors reported ambitious publication plans. Over 85% reported that they planned to publish articles, proceedings, chapters, books, and other works based on their theses and dissertations. Only 13.7% reported that they did not plan to publish anything from their ETDs. Over two-thirds of our graduate students reported that they would publish articles, and 40.7% anticipated publishing in conference proceedings. Only a small number, 26 out of 445 respondents or 5.8%, plan to publish a chapter, while a few students, 31 or 7%, plan to publish a book.

While many authors and their faculty advisors are afraid that ETDs that are available on the Web will be considered "published," only 11.4% of the authors reported that publishers had advised them to restrict access to their ETDs. On the

other hand, the graduate student authors reported that their faculty are overwhelmingly advising them to restrict access to their ETDs—235 out of 271 or 86.7%. Someone other than faculty and publishers advised 82 authors (30.3%) to restrict access to their ETDs.

ALUMNI'S PERSPECTIVES

In addition to having information from graduate students who submitted ETDs in 1999, the Graduate School gathered data from about 50 alumni who allowed their ETDs to be Web-accessible during the previous year. While 86.3% of the 1999 ETD authors said they would publish something from their ETDs, 42.6% of the 1998 survey respondents said they had published. Remarkably, the published authors reported that they did not encounter any resistance from publishers to accepting their manuscripts because they were derived from online theses or dissertations. Hopefully the same will hold true for the 1999 ETD authors though 11.4% said they restricted access based on advice from publishers.

The 1998 authors responding to the Graduate School's survey reported overwhelmingly that they were satisfied (66.7%) or somewhat satisfied (29.2%) that their work was more widely known and appreciated because their ETDs were accessible. Over 40% had been contacted as a result of having their works on the Web, and 88.3% were satisfied or somewhat satisfied with contacts resulting from their Web-accessible ETDs. The vast majority were satisfied (27.8%) or somewhat satisfied (55.6%) that it also helped to expand their network of research colleagues. Of those who received comments about their ETDs, 83.3% reported receiving somewhat positive comments though 16.7% said they received somewhat negative comments. Small, but equal numbers were somewhat satisfied (5) and somewhat unsatisfied (5) that having their ETDs online helped them locate employment.

Among the comments they noted that they were pleased with the ability to "share one's knowledge and research." Another popular comment was about how "incredibly convenient" it was to provide copies of their theses and dissertations by forwarding the URLs of their VT ETDs. One survey respondent expressed concern about the ease of plagiarism and copyright violation. The Graduate School is conducting another alumni survey in September 1999.

USERS' PERSPECTIVES

In addition to survey responses from ETD authors, we collected responses from about 50 individuals between April and October 1998. Voluntarily completing a "user survey" that was a link from the library's ETD homepage (<http://scholar.lib.vt.edu/theses>), they responded to 16 multiple-choice questions. Where did these respondents work or study? 65.4% were associated with universities, 13.5% with industry, 7.7% with schools, and 13.5% worked or studied elsewhere. Who were these respondents? 44.2% were researchers, 3.8% were faculty, 11.5% were teachers, and one (1.9%) was a librarian. 13.5% categorized themselves as "other."

Of our 52 respondents, nearly one-third (32%) had relatively fast Internet connections (i.e., Ethernet, T1, ISDN, or cable modem). The majority, 54%, was connecting at slower speeds, 56 KBPS (kilobytes per second) down to less than 14.4 KBPS. The connect speed was unknown to seven of our respondents.

These survey respondents used PCs, Macs, and Unix-based operating systems in similar proportions to ETD authors. 2.4% of the user survey respondents used Unix-based systems. 13.7% of the authors used Macs, but slightly fewer of the surveyed users did, 10.2%. ETD authors were 83.7% PC users and 91.8% of the surveyed users. Only about one-fourth of the ETD users responded that they were unfamiliar with PDF.

Twenty-five percent of our ETD users were from universities that do not accept ETDs, though 27.5% responded that they did not know if their universities accepted ETDs. 40% were from universities that accepted ETDs. 18% of the respondents said they submitted ETDs, while 82% (32) said they did not. 55.2% of the user survey respondents said their universities *should* accept ETDs. Only one person answered that their university should not accept ETDs, and 20.7% had no opinion on this topic.

When asked why they came to the ETD digital library homepage (<http://scholar.lib.vt.edu/theses>), the largest percentage, 62%, said they were doing research. 21% wanted to learn about ETDs, 2% were there because of their jobs. Personal interest was the motivating factor for 15%. There was an even distribution for responses to "Were ETDs easy to find?" Seven responded "very easy" while six responded "very difficult." Ten each responded easy, fair, and difficult. Four responded to our survey before they tried to find any ETDs. In the future we may

survey most of these readers again. Over 82% said that we could contact them for further information and they gave us their email addresses.

FACULTY MEMBERS' PERSPECTIVES

ETDs have been a topic of conversation among Virginia Tech faculty for many years and in the winter of 1996 the Degree Requirements, Standards, Criteria, and Academic Policies Committee (DRSCAP) recommended to the Commission on Graduate Studies and Policies (CGS&P) that graduate students submit theses and dissertations electronically. Unlike the information we have gathered from student and user surveys, faculty opinions have been gleaned largely from electronic mail archive (email), minutes of the university's ETD Advisory Committee meetings, and anecdotal evidence from workshops and individual encounters.

One source is the archive of minutes of the Commission on Graduate Studies and Policies and its Degree Requirements, Standards, Criteria, and Academic Policies Committee. The minutes of October 18, 1995 show a proposal from DRSCAP that recommends that the Graduate School begin requiring students to submit their theses and dissertations electronically by the fall 1996 semester. The minutes of a January 1996 Commission meeting report six concerns: (1) phase in the requirement, (2) make resources available to support the requirement, (3) include copyright statement, (4) put in place procedures for exemptions, and (5) make allowances for hard copies and (6) adaptation to new software.

On Feb. 21, 1996 CGS&P received a document on implementing ETDs prepared by DRSCAP. After modification, it was approved unanimously. One of the modifications was to begin the requirement Jan. 1, 1997. DRSCAP recommended: (1) encourage voluntary submission prior to January 1997; (2) establish university oversight committee; (3) publicize the requirement widely; (4) provide training and workshops for students, faculty, and staff; (5) avoid significant burden on departments.

At the April 17, 1996 meeting of the Commission, an amendment was made to the Feb. 21 policy stating that ETDs would not be available on the WWW until a release form granting permission to do so was signed by both the student and major professor. The oversight committee was also charged to work with societies and publishers to encourage them to adapt their policies concerning ETDs.

As a member of the Virginia Tech ETD Team, my email archives are quite robust. They show areas of faculty concerns focussed on three main topics in 1996: (1) quality of the works representing the university; (2) easy and timely access to graduate students' research; (3) impact of ETDs on future publications.

Of these concerns, only the last was reflected in the minutes of DRACAP and CGS&P. Widely available ETDs caused faculty to be more concerned about the potential impact of their students' works. They wanted their university to be represented by quality research, and, if they are going to be available, they should be timely and easy to find.

Many email messages also focused on training for students and faculty. This concern was addressed in two ways. The Graduate School has provided "ETD Workshops," informing attendees about the process of conversion from word processed documents to PDF and the online submission process. At the workshops, students frequently ask about scanning images--an indication of their planning to convert analog formats to digital formats instead of creating digital works initially.

Secondly, Faculty Development Institutes (FDI: see Endnote) introduced faculty to similar processes. In 1996, 340 Virginia Tech faculty attended week-long FDIs, subsequently receiving the hardware and software that they had been instructed in using. Seven-hundred faculty had participated in the faculty development initiative by the end of the first year of the ETD requirement.

The third issue of concern to faculty in 1996 was the impact of future publications and whether an ETD would be considered published if it was available on the Worldwide Web. Unable to definitively answer this for every publisher, Virginia Tech responded to this concern by offering restricted access options. This was to give publishers time to become familiar with ETDs and reduce the perceived threat that ETDs posed journal sales. In addition to the option of unrestricted, worldwide access, students and their advisors have the option to limit access to the university community or to withhold access from everyone, both for a limited amount of time.

In 1997 the issues of concern to Virginia Tech's faculty are documented in articulate, prolific, and heated email transmissions. Many faculty felt the ETD requirement had been handed down to them by the university administration without any opportunities for faculty input. Evidently, faculty on the CGS&P and DRSCAP had not taken information about the ETD initiative back to their constituencies. In

addition, few faculty had read the articles published in the campus newspapers. The faculty, staff, graduate student newspaper, *Spectrum*, mentioned electronic theses and dissertations almost monthly in 1996, including workshop announcements, reporting an article about the Virginia Tech ETD initiative in the *Chronicle of Higher Education*, recounting meetings of the CGS&P and DRSCAP, announcing a FIPSE grant (Fund for the Improvement of Post-Secondary Education), and more. In 1997 ETDs were mentioned bimonthly, including articles about the feature story on National Public Radio, Microsoft and Adobe software donations to the NDLTD, workshop announcements, and more.

From the email archive four themes important to the faculty in 1997 stand out: potential harm to future scholars, "cultural" differences among academic disciplines, faculty feeling threatened, and the lack of clarity about intellectual property law. Only the final issue continued to be of concern to faculty in 1998 and 1999.

The emails stressing the potential harm to future scholars had three components. (1) Would new scholars be able to get their articles, for example, accepted for publication if ETDs were considered published because they were Web-accessible? (2) Would the online works have the same quality as the earlier paper versions? The subtext of this concern was: (a) Would ETD authors continue to speculate about future research if they were afraid that their ideas could easily be plagiarized? (b) Broken Web links would detract from the work, and (c) big files would be too slow for readers to download. The last component (3) was how would UMI handle ETDs in *Dissertation Abstracts*?

Getting ETDs into *Dissertation Abstracts* was an early workflow issue resolved through programmatic email notification to UMI of each approved and available VT ETD. However, fear for students whose ETDs would be fully Web-accessible lead over 62% of our faculty to advise their students to restrict access, according to authors surveyed in 1999. 20% of our students have responded to this cautionary advice by withholding all access, and 30% restrict access to Virginia Tech. This advice has also lead to *Dissertation Abstracts* listing fewer Virginia Tech ETDs. Unlimited, worldwide access is available to 50% of the ETDs approved by the Graduate School.

Rating the quality of ETDs in comparison to other formats has not yet been address and is a topic for further study. ETDs are richer in graphics and color, very

likely due to the reduced expense producing the online version, versus the considerable cost of supplying multiple color copies that would be required for the paper format. The quality of ETDs with broken Web links has not been studied. However, many feel that an out of print article is also very difficult to find and probably leaves the researcher with a greater dearth of information than do broken links and missing Web sites. The relationship of quality to download time is questionable, especially when heavy Internet traffic slows access, but it is certainly faster than traveling to campus and parking, or waiting for a dissertation to arrive through document delivery services such as interlibrary loan or UMI.

LIBRARY PERSPECTIVES

The advent of electronic theses and dissertations also raised a cadre of issues and responsibilities for libraries to address, including improving access to information, maintaining the information server, and archiving. With ETDs libraries became better-stocked digital libraries, provided more timely access to information, served more users without increased staffing, and saved shelf space. Some issues are new to libraries but most require mapping traditional services and resources to digital resources and electronic services.

The library is a traditional source of information, but it has not necessarily established itself as the locale for computer training. This is changing, not because of, but along with, the advent of ETDs. There are typically two types of library computer labs: those supporting learning about access to resources and those supporting training and resource development. Some libraries provide both, assisting authors to prepare ETDs and introducing users to new information resources such as ETDs. According to the author surveys, library instruction has been somewhat effective based on author use of the New Media Center. Faculty also receive training in the use of technology to prepare electronic works for publications and they are introduced to the library's new information resources.

Librarians with increased instructional responsibilities assist users throughout the full information cycle, from discovering to communication to submission, possibly for online access and publication. In addition, computer labs in libraries can familiarize graduate students with asynchronous collaboration tools, enabling them to work with local as well as remote resources. As graduate students become more

sophisticated software users (as are today's incoming university freshmen), libraries may provide authoring spaces with less intervention.

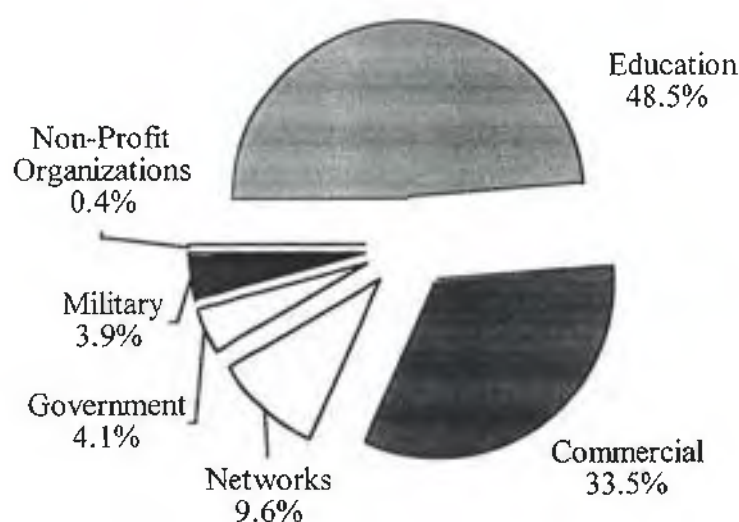
With ETDs libraries enable universities to unlock intellectual property. Unlike the research described in printed and bound theses and dissertations that has been essentially hidden because it circulates so infrequently from library shelves. Similarly, very few dissertations, available through *Dissertation Abstracts*, have been requested frequently enough to garner royalty payments. Virginia Tech's ETDs are available to a wide community of users unhampered by the hours the library is open and this results in far more frequent use than the paper or microfilm formats.

ETDs Requested			
<u>1996</u>	<u>1997</u>	<u>1998</u>	<u>Jan.-Aug. 1999</u>
4,600	72,854	244,987	476,313

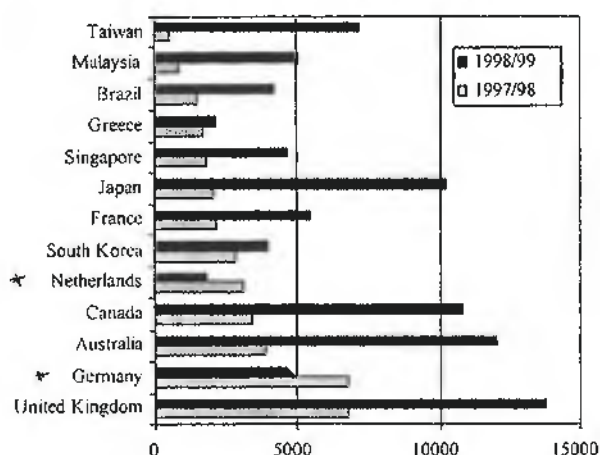
Computer log files reveal that not only does the library provide increased access to its ETDs, its clientele is broader than the local educational domain and now reaches commercial and government domains.

ETDs are labor saving devices! Increased use does not require increased library staff time because ETDs do not have to be mailed to *Dissertation Abstracts*, bound, labeled, security stripped, barcoded, checked out and checked-in, shelved and reshelved.

With ETDs, Libraries Serve More Users in the United States



With ETDs, Libraries Serve More International Users



Computer programs "move" submitted works through the approval process to availability and into the archive. Even without deriving the cataloging record from the online text, eliminating handling of multiple paper copies of each title could save 73.3% of the cost of processing paper dissertations.

Libraries continue to provide access to ETDs from within their buildings for patrons. At many research universities, libraries also often provide "better" downloading times through Ethernet connections. Therefore, libraries provide Internet workstations throughout their systems for students to access completed ETDs as well as supporting computer labs for authors to prepare these works.

In the future, libraries may be the principal locale for the variety of software necessary to view multimedia ETDs. Standards for software used by ETD authors are developing by default, rather than through mandates or requirements.

Multimedia Recommended / Used in ETDs

<u>Formats Recommended</u>	<u>Formats Used</u>
IMAGE: bmp, dxf, gif, jpg, tiff	.dxf, .gif, .jpeg, .pdf, .tiff
MOVIE: avi, mov, mpg, qt	.mpeg, .qt, .eps
SOUND: aiff, mcd, wav, mp2	.aif, .wav
TEXT: pdf, html, SGML	.pdf, ETD-ML
OTHER: Macromedia, SGML, XML	Authorware, Director, xls

Some library users are uncomfortable when libraries do not have copies of ETDs in paper, preferably, but also microfilm formats. There are two important factors, however, that many library users do not consider. One is that some ETDs were not designed to convey the same message as text on paper (see the list of media above). Over time users will become accustomed to this media, especially when they enjoy, for example, the benefits of online works having multiple simultaneous users. Advantages also include the end of mutilated, destroyed, or stolen materials requiring expensive replacements or elimination of the works from library collections. The security of multiple copies of every electronic work should alleviate some of the tension created by not having these works also available on the shelves.

ETDs can become accessible immediately, once the appropriate university unit completes the evaluation of the work. There do not have to be any processing delays as with paper. In addition, a library's database of ETDs can be reindexed frequently so users can reliably search and retrieve newly available works.

Quality as well as timeliness of cataloging records for theses and dissertations has improved. When the bibliographic information from the title page is available online, catalogs can provide easy access to this information easily. When the abstract is also online, more information can be included in the catalog's bibliographic record. Library systems eventually will programmatically derive the cataloging records directly from the digital works, and with more computing power, some institutions will index entire ETDs and every word in these works will be a point of discovery for Internet users.

While libraries improve workflow and take advantage of the ease of providing prompt access to ETDs, they also have the continuing responsibility to maintain long term access and to archive these works in their final form. Maintaining the server for ETDs need not be a huge added responsibility for libraries that already maintain computers for storage and access to other digital works. Libraries, especially those that have a tradition of strong internal systems support, also have a responsibility to preserve ETDs at the same time that they provide continuous customer support.

One of the best security practices is for multiple agencies to reciprocate online archiving, and the most effective way of doing this is through "mirroring." While this has as yet not been formalized, it is a concept under discussion among members of the Networked Digital Library of Theses and Dissertations (NDLTD). But well managed, copies can also effectively back-up online resources. Many academic libraries are working with their computer science departments where research and development in digital libraries holds great promise for the future. But even so, the fact that PDF is a pervasive format used heavily by the federal government as well as many commercial publishers, means that university libraries will not be alone in trying to solve migration issues.

CONCLUSION

Graduate students who develop ETDs, will be better prepared to enter and fully participate in academia in the Digital Age and to be more effective digital library

users even outside of the academy. They will be accustomed to creating online resources in addition to finding and using them. As novice academics they will be better prepared to submit articles electronically if they have already learned to develop and submit their digital graduate research. Libraries can be the locale where the full cycle of information takes place, from creation to access.

The speed with which research becomes available, combined with easy access on the Internet and Web, has resulted in many graduate works getting the exposure and use they deserve. The scholarly community, both here and internationally, as well as American government and commercial users, welcome timely and free access to the research available in ETDs.

Nearly 60 universities throughout the United States and internationally (see <http://scholar.lib.vt.edu/theses/NDLTD/members.html>) are participating in the evolution of electronic theses and dissertations by joining the Networked Digital Library of Theses and Dissertations. They are adapting traditional procedures and workflow to incorporate local authors' direct submissions into timely, worldwide access. The computer programs and scripts written by the Digital Library and Archives staff to process ETDs are available without charge. While many are anxious about the unanswered questions remaining (such as the durability of digital works), many more are satisfied that ETDs receive increased use instead of adding to the thousands of theses and dissertations gathering dust from lack of use in university archives.

Through the perspectives of students, faculty, readers, and libraries, this paper has reviewed some of the immediate, near term, and long term issues related to electronic theses and dissertations. At Virginia Tech we have garnered experience from over five years of direct contact with many cultures, both within and outside the academy, and they have contributed to the evolution of theses and dissertations from paper to digital formats and their dissemination throughout the world.

DATABASE STUDIES

Methodology

The first attempt at a study of the transportation literature reviewed what types of materials were in the appropriate databases. This would be the first resource used by a librarian for a literature review and an excellent research tool for the practicing transportation researcher. Many projects would start in this manner. Once a listing of pertinent resources was retrieved, the availability of these resources would be of vital interest. Items in a database might produce an interesting bibliography but the primary resource is essential for quality research. Therefore, not only were the types of resources indexed in the databases important, but the ability to actually obtain the resource in a timely, cost effective manner was of interest.

To determine availability of the records indexed in the database a sample of database records was collected in 1995 and another in 1996. The methodology of each study was identical. Each took a random sample of records for the same topics and years, coded the record for format, language, and country of origin for reports. Next step was to search each record in the Online Computer Library Center (OCLC) for holding information. OCLC is a library utility that houses records of resources from over 30,000 libraries from around the world. Current statistics from the company's web site state that more than 38,000,000 records in over 400 languages are in the database which also includes a listing of libraries that own an item. It is a major source for locating libraries from which to request an interlibrary loan.

The first study was based on the TRIS transportation database from the Transportation Research Board, at that time one of the major transportation databases and widely available through the DIALOG Information Services group of databases. Two hundred seventy eight records resulted from the subject searches. The second study was performed in TRANSPORT, a new database that included TRIS and the International Road Research Documentation database from the Organization for Economic Co-operation and Development (OECD) and the European Conference of Ministers of Transport's TRANSDOC database. The subject searches in TRANSPORT resulted in 540 records.

Results

Much of the study covered what could best be described as the basic demographics of the database records. For example, the format types were tabulated for both the 1995 and 1996 studies and are summarized in Table I. English was the major language representing 91% of the first study and 83.6% of the second.

Take in Table I

The Luxembourg definition which states grey literature is "that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers" was used. (GreyNet Website) In the two studies, 73% and 85% of the citations were to grey literature.

However, it was the OCLC results that were the crux of the study. OCLC represents the largest, most commonly used resource for the sharing of resources in the U.S. and Canada and does have a broad international perspective. The most important question of the study was whether citations from the database searches were actually available in the U.S. and Canada for fast and cost effective use. OCLC provided the means to answer this question from an American perspective.

The results of both studies were similarly grim. Overall, 77% and 76.7% of the records were held by at least one U.S. or Canadian library. Initially that might appear positive but in reality roughly one in four of the records was not readily available, not an overwhelmingly encouraging number.

However, when the records are segregated by language, a more disturbing trend becomes evident. Zero holding locations for the U.S. and Canada now represent 64% and 71% of the records. Additionally, 8% and 17% are only held by one location.

Implications were clear: on the one hand, the best information available was needed to advance transportation studies into the next century and this included many resources from international resources, but on the other hand, indications were that a number of resources were not readily available. Research and observation have indicated that timely access is crucial in technology, yet many resources were not readily available to the researches.

BIBLIOGRAPHY STUDY

Methodology

The next step was to expand the study and go beyond the records indexed in a database to the actual literature. What were the transportation researchers citing in their published papers? Were they able to obtain the resources they needed, regardless of language and location of publication? Did their actual references mirror the databases or did they more closely reflect national holdings?

To provide a snapshot of this aspect of the transportation information, eleven papers from various publications from the Transportation Research Board (TRB) and FHWA were analyzed. The eleven included papers from the *Transportation Research Record* and the *Transportation Research Circular* and reports published by the FHWA. The papers were published during 1997 and analyzed in December 1997. A total of 367 citations were included in this study.

The results can be summarized as follows. Books and conferences represented 39.3%, journal articles 30.7%, reports 28.7% and theses 1.3%. English language publications totaled 96% of the references. In the report format, only 12.5% of the citations were to reports from outside the U.S. Grey publications totaled 86% of the references.

It was evident from these snapshots, taken from a number of vantage points, that there was indication of a potential problem with obtaining and using international transportation information with inherent implications that were highlighted in the *Value of Information* report and other reports on this topic. It also was becoming evident that the U.S. was not collecting international resources at a significant level and the researchers were not citing these resources in their publications. These were important issues that several organizations, including TRB, the U.S. Department of Transportation's Federal Highway Administration, the American Association of State Highway and Transportation Officials (AASHTO) and the Special Libraries Association's Transportation Division, were considering.

However, several other questions were raised from these studies. Were our international colleagues facing the same problems? What resources did the researchers abroad use and were these resources available to U.S. and Canadian researchers?

With the support of the FHWA and AASHTO, a study of eleven transportation centers in nine countries was conducted during the second half of 1998. These centers represent most of the major European centers. Several developing countries which have relationships to U.S. and other European transportation centers and who utilize technology transfer centers in the U.S. and Europe rather extensively were also included in the study. They could provide an indication of information needs, utilization and production of information from a different perspective.

The transportation centers visited included: Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland; Estonian National Road Administration, Tallinn, Estonia; Federal Highway Research Institute (BAST), Bergisch Gladbach, Germany; Finnish National Road Administration (FinnRa), Helsinki Finland; Swedish National Road Administration (SNRA), Borlange, Sweden; Swedish Road and Transport Research Institute (VTI), Linkoping, Sweden; Technical University of Delft (TUD), Delft, The Netherlands; TNO-Delft, Delft, The Netherlands (a for-profit agency); Transport and Road Research Institute, Kuanas, Lithuania; Transport Research Centre, Brno, Czech Republic; Transportation Research Laboratory (TRL), Crowthorne, The United Kingdom.

A list of questions was generated and reviewed to gather appropriate demographic information. A random sample of locally produced reports was selected and the bibliographies and title pages were copied for analysis upon return to the U.S. References were collected from EPFL, BAST, FinnRa, SNRA, VTI, several Dutch federal agencies, and TRL. Reports were not available for study from the Czech Republic, Estonia or Lithuania.

Results

During the interviews all interviewees stated that most of the center's researchers were conversant with at least one language beyond the native language. In many cases, the researchers were comfortable with several languages. English was widely read and/or spoken and English language materials were not a problem for most of the researchers. In addition, some level of expertise in many languages was available at the centers. Therefore, translation assistance was frequently in the center itself, at least to the level needed to translate needed graphs and tables. While translation services generally were used when needed, they were not used

frequently due to the in-house capabilities. This contrasts significantly with the more mono-lingual U.S. research population.

The citation analysis was based on the collection of reports collected from the centers and U.S. libraries. A total of 35 reports were collected from Finland, Germany, The Netherlands, Sweden, Switzerland and the United Kingdom yielding a total of 627 references.

Format types are summarized in Table II

Take in Table II

The composite format percentages are reports (58%), books and conferences (26%), journals (12%), theses (1%) and others (3%). The 'other' format type includes web pages, brochures and catalogs.

In most cases native language resources represented the greatest percent of the total. The Netherlands was the exception. The percent of native language to total citation numbers was Finland 78%, Germany 68%, The Netherlands 29%, Sweden 84%, Switzerland 97% and the United Kingdom 93%.

The number of languages represented in the cited references ranged from two for Germany (German and English) to seven for Finland (Danish, Dutch, English, Finnish, German, Norwegian and Swedish). Language representation as a percentage of the total was English (36%), Swedish (19%), German (19%), Finnish (14%), French (6%), Dutch (4%), Italian (2%), and Danish and Norwegian (each .4%).

The last analysis of the report references was a determination of the availability of each resource in the U.S. or Canada. Each citation was searched in OCLC and the number of U.S. and Canadian libraries that were listed as lenders was noted. Results are summarized in Table III.

Take in Table III

DISCUSSION

The findings of this study of transportation bibliographies, published by a number of international centers, mirror the snapshot of bibliographies from TRB publications. First, centers have a very strong tendency to cite resources in their own languages. In many cases this translates to locally published reports. Although all interviews indicated that the local researchers read several languages, the native language was almost always the most often used. In this study, non-native language

citations rate was 7% for the United Kingdom, 16% for Sweden, 22% for Finland, 32% for Germany and 71% for the Netherlands. Switzerland had only 3% of the citations in non-native language, but since French, German and Italian are official languages, this figure is somewhat misleading. Still, the majority of citations were to native language publications, negating some of the benefits of superior language skills.

The second most common language is English and represented 36% of total citations. Most collections had a number of U.S. publications in their collections, most notably U.S. Department of Transportation reports and the various publications of the Transportation Research Board.

Second commonality is the preponderance of grey resources. Publications from commercial producers are included in the citation analysis. Most of these citations are to the major commercial journal publishers. However, the majority of citations fall within the definition of grey literature written at the Luxembourg conference in 1997. In the earlier database reviews, 73% of the records in the TRIS study and 85% in the TRANSPORT study were to grey literature. In the TRB/FHWA bibliographic reference analysis the number was 86%. The results from the references cited in the international reports are similar to the percentages found in the similar, with 89% of the citations to grey literature. Table IV summarizes the percent by country. # GL

Take in Table IV

While there is some variation between the studies, most notably the TRIS study, the other three hover at 85% or more references to grey literature. This is a high number and is likely one factor in the difficulty in obtaining transportation literature. An analysis of reports by specific country was not computed, however, a cursory glance at the report literature indicates that the native reports and the U.S. reports are the most popular, indicting the prevalence of local reports and those from the U.S. and a relative dearth of non-local reports. Even those countries that cited a number of nations tended to use those from surrounding countries.

One issue that was made clear in the interviews was a different point of view concerning free access to information and the use of the World Wide Web. Both are linked since they are linked to the issue of fiscal responsibility.

Most of the transportation centers are responsible to some degree for income generation or cost recovery. The sale of their publications is one means to generate income. While they may be willing to link their publication list from their web site,

they are not able to link the full text document for free access. Instead, reports that might have been exchanged with other libraries or sent gratis are now sold at significant cost to the buyer.

In a similar manner, the Web is not viewed with the same enthusiasm by some of the European centers. The Web sites are created for promotion of the centers and their products, not free access to the products themselves. Their publications are copyrighted and generate income for their facility. This stands in opposition to the U.S. trend of placing full text documents on local web sites or from the more comprehensive National Transportation Library (NTL) web site maintained by the U.S. Department of Transportation's Bureau of Transportation Statistics. Thousands of reports, dozens of databases and directories are freely available in a searchable web site.

Any hope that similar virtual libraries would be created by the international centers faded during the interviews. There is little interest in loading full text reports and databases for free access. Since most of the reports are copyrighted, they cannot be mounted by the NTL. Therefore, access will continue to be through traditional collection development routes.

Collection development has been stressed by increasing demands, increasing number of resources published, increasing prices and decreasing budgets and limited space. However, in transportation this is also linked to the large number of resources that are grey literature. [While grey literature is far from impossible to collect, it does present another level of complexity to the already difficult collection picture]. More items need to be searched and ordered individually and some of the resources are difficult to locate through routine procedures. This is one of the reasons the anecdotal information stated that so many resources could not be located for loan or purchase.

The general picture from the interviews and the analysis of database records and cited references was that there is a wealth of valuable resources to support research and improve transportation studies. However, that wealth of resources is closely linked to a poverty of access points for researchers to obtain the resource, whether in print or in electronic format.

RECOMMENDATIONS

While the picture that is painted by this multi-faceted study is rather dismal, there are several hopeful signs. First, there is a great deal of interest by various organizations to formulate a solution and devise and implement an action plan. The National Cooperative Highway Research Program (NCHRP) has funded a proposal to investigate the collection, storage, dissemination, marketing and translation of non-English transportation resources. It will initially focus on French, German and Spanish language resources but could be expanded at a later date.

While this process is underway, other initiatives are needed. A large, formal strategy is recommended, there are more 'grassroots' actions that can make a significant difference in a short time.

First is the need to know what is really available in the Americas. There are numerous collections in state departments of transportation, specialized academic collections and transportation organizations that are not cataloged on the web. The simple loading of a library catalog or a collection list could provide the exact information needed. These could be searchable through the search engine on the NTL, with a link to borrowing information. There are a significant number of libraries that are not and may never be listed in OCLC and their collections are known to a few. This is a low cost solution to a basic problem. The benefit is free access to the catalog record for all with Web access.

Second, the complexity and need for a systematic collection of international resources needs to be acknowledged and addressed. There are still opportunities for exchange of materials. While there are probably few U.S. institutions that can afford the exchange costs or storage costs of a exhaustive collection, some can take on the responsibility of selective language, country or topic collection and widely share this information. There is the possibility that some of this may be funded by a coordinated national strategy recommended through the NCHRP project.

The statistics indicate the importance of grey literature to the field of transportation. Locally produced reports are heavily cited in bibliographies. The relatively ubiquitous U.S. Department of Transportation reports are well represented also. This is a strong indication of the importance of this very primary source of research information, yet there is a definite difficulty in obtaining non-U.S. reports in the U.S., and to some extent other countries. A wealth of technological innovations,

new techniques and research dead end warnings are going unheeded due to lack of access to the literature.

Increased attention must be paid to the non-commercial aspect of the transportation literature. While the major organizations and government publications are relatively well known and collected, materials of a more local nature and other publishers need to be monitored and collected. Items such as brochures and catalogs can provide valuable information, but are not always kept, and if kept, may not be cataloged.

Third, while there are realistic limitations to the use of the Web, international co-operation should be pursued. There are some items that can be made available and even catalogs and lists that are maintained are informative.

In a similar manner, the inclusion of as many resources as possible in the transportation databases should be encouraged. The wider the exposure of the resources, the greater the likelihood that someone will want the item and it will be included in a library collection and available for loan.

CONCLUSION

Transportation literature relies heavily on report information. Since these reports often are, by the nature of the funding and utilization patterns, the product of government agencies, they are in the accepted definition of grey literature. Many people have studied and described the intricacies of grey literature. Many have emphasized the vital importance of this type of information. Few fields have such a high percentage of information in the non-commercial arena. Transportation presents an interesting, intricate and important subject to develop efficient and systematic procedures for the identification, acquisition and marketing of grey literature. Attention to this process may be critical in efficient progress in the field of transportation.

Table I Format Types

Format	1995 Study (TRIS)	1996 Study (TRANSPORT)
Journals	28.4%	22.2%
Books/Conferences	21.6%	41.7%
Reports	49.6%	35.7%
Theses	.4%	.4%

Table II Format Type by Country

	Finland	Germany	The Netherlands	Sweden	Switzerland	United Kingdom
Reports	31	33	36	88	55	64
Books/Conf.	52	19	21	10	7	27
Journal articles	10	62	18	6	10	12
Journals (unique titles)	7	27	7	6	8	10
Theses	0	0	0	1	2	3
Other :	0	9	1	4	3	1

W/B
Page
Brochures
Catalog

IN USA OR CANADA

Table III Number of Holding Locations as a Percentage of Citations

	0	1	2-5	6-10	11-25	26+
Number of Citations	379	38	32	12	19	47
Percentage	72	7	6	2	4	9

Table IV Percent of Grey Literature Cited in Locally Produced Reports

Finland	Germany	The Netherlands	Sweden	Switzerland	United Kingdom
91%	80%	90%	97%	84%	91%

as diverse as health, engineering, chemical and environment. The subject of safety also constitutes a vast and diversified field of research for both research centres and institutions operating in the sector.

The importance attached to information has resulted in the production of a great quantity of documents, most of which share the typical characteristics of Grey Literature (GL). The retrieval and management of such documents is, however, critical due to their limited availability, the difficulty in pinpointing document producers and the lack of bibliographies gathering information on this type of documentation.

This paper illustrates the results of the first phase of a two-year project funded by the Italian National Institute of Occupational Safety and Health (ISPESL), the technical-scientific body of Italy's Health Ministry, which is fully involved in the promotion of new initiatives in the sphere of prevention and safety in the workplace. The main aim of the project is to develop a database to be distributed on information networks, containing Italian GL in the field of occupational health and safety.

The design of a data base entails not only the selection of the best technological solution, but above all decisions on the type of information that the database should contain, acquisition methods, the organisation and management of documentation as well as an analysis of actual and potential users. Due to the inherent characteristics of GL the data acquisition phase is critical: producers must be identified along with their methods of publication and distribution, even if these are limited and informal [5, 6].

The first phase of the project therefore consisted of a questionnaire-based survey aimed at potential GL producers, in order to identify sources, types of documents produced, publication, dissemination and bibliographical treatment.

The survey also has the secondary aim of finding out whether special types of GL documents are produced in the health and safety sphere which contain information specific to this sector.

This paper is divided into three parts. The first part describes the survey's methodology, the second analyses the main results from the various sections of the questionnaire and the final part examines in detail and dissemination modalities.

2. Methodology

GL production generally reflects the activity of its producer, and indeed a distinction is often made [7, 8, 9] between GL produced by research centres or by government agencies or the private sector. GL is moreover produced both to *activate* communication within the same institution and to communicate with other organisations [10]. These distinctions reflect the production of special GL document types which indicate organisations covering specific roles and document types which have their own contents and objectives. This reveals the close relationship between GL production and its producer, the identification of which is the first, but essential step on the way to identify, gathering and managing these types of documents.

The difficulty in identifying the GL producer is compounded by inherent difficulties in the occupational health and safety sector in general, i.e. the number and diversity of subjects involved, the tasks and duties performed by each subject and the application of different disciplinary field. The conceptual scheme (Fig. 1) illustrates the various elements to be

Said of each
discipline/phase

considered in the analysis of this field and their main relationships. This scheme also seeks to highlight the different points of view from which the sector can be analysed, and has been used to select the survey targets.

This sector is extremely wide, due to the fact that health and safety involves the entire production system (public administration, firms producing goods and services, etc.), in the public and private sectors (central and local administrations, private firms, etc.), whatever their size and whatever the economic sector (manufacturing, farming, etc.). Health and safety activity also involves all subjects in the production system which have to conform to the law as well as institutions with monitoring roles. Of particular importance among the latter group are those organisations and/or services set up when the current law came into force.

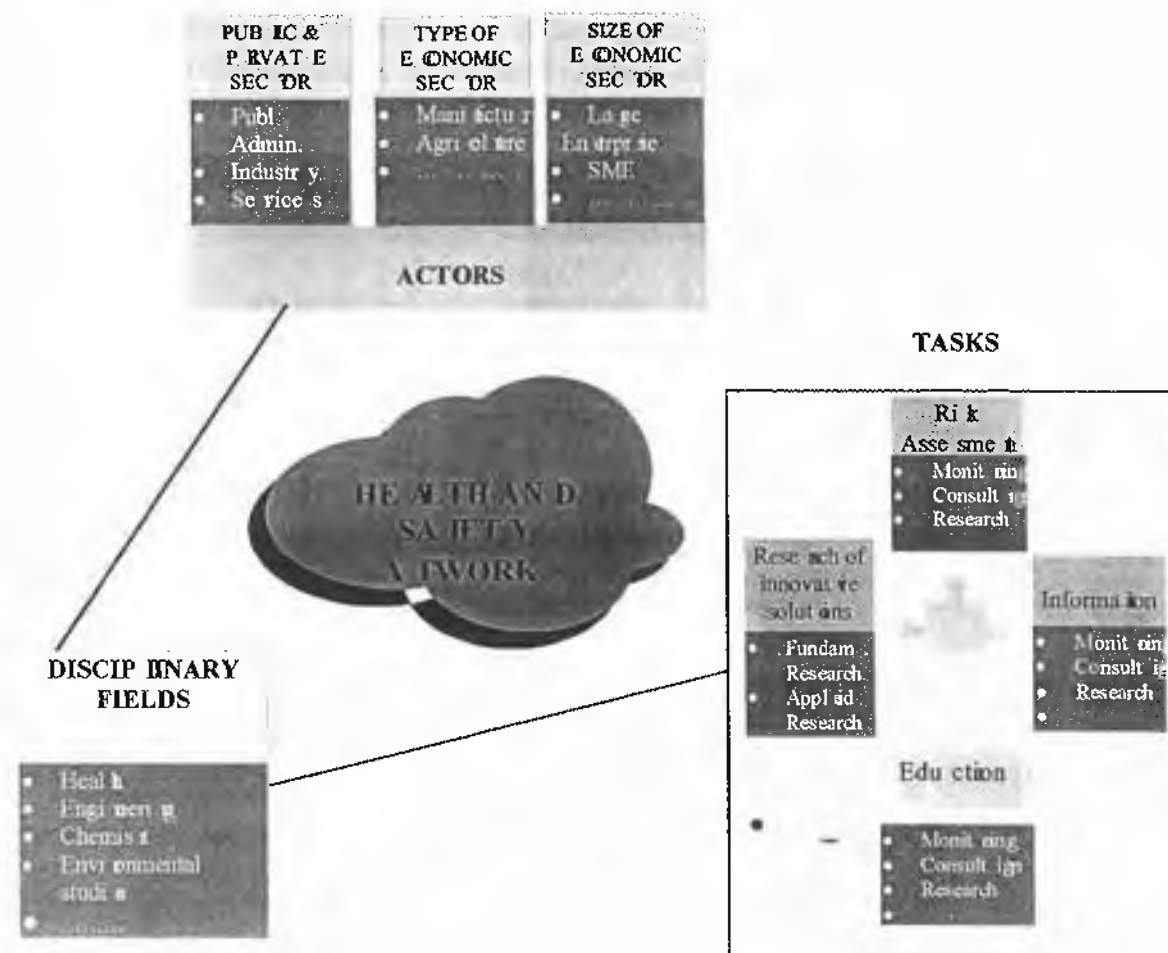


Fig. 1 - The main elements interacting in the field of health and safety

The survey on GL producers gave priority to the public sector, while the variables of the productive sector and that of organisation size were not taken into account.

The health and safety sector is multidisciplinary, ranging from the health to environmental issues. Generally speaking, however, since the law sets out to protect the health and safety of workers in any working environment, the most commonly used disciplinary competencies are medical, to safeguard workers' health, and engineering, for the correct use of equipment. This is why those organisations operating in the sphere of prevention in the medical and engineering fields were selected.

Summing up, the area at which the survey was aimed was selected on the basis of the institutional roles of organisations and/or services and the main research, monitoring and consulting activities performed by them. Priority was also given to the public sector and to the medical and engineering areas. The main criterion adopted was that of the representative nature of the sector and that of uniformity in relation to the type of organisation to be chosen for the survey.



Fig. 2 - Surveyed organisations broken down by function

Of those organisations chiefly dedicated to research (Fig. 2) we have selected, from the faculties of medicine and engineering, the University institutes most closely involved with health and safety (Industrial Medicine Institutes, Hygiene and Preventive Medicine Institutes, Forensic Medicine and Insurance Institutes, Departments of Building, Construction Sciences, Material Sciences, Mechanical Engineering, Chemical Engineering, Electrical Engineering). Similar criteria were adopted when selecting CNR (National Research Council) Institutes. With reference to ISPESL, the questionnaire was sent to all of its peripheral Departments, which operate closely with local areas. Finally, the grouping "other Research Institutions" included both public and private institutions.

Organisations which carry out monitoring activities include regional authorities (Health Departments Prevention Services and Regional Environment Agencies) and Local Health Units (LHUs). Within the LHUs we have selected both the Internal Occupational and Health (OSH) Responsibles and the External OSH Services. The first are charged with the task of ensuring that the law is applied within their own institutions, while external OSH Services provide guidance and ensure that laws are complied with at a local level.

Organisations which carry out consulting activities include Trade unions and Employers' Associations. Since they represent workers and employers, they often promote initiatives to disseminate information about occupational health and safety. This grouping also includes the associations of special professional categories expressly dedicated to prevention (National Society of Prevention Workers, Italian Inter-Association Prevention Council, etc.).

To identify GL producers in such a heterogeneous environment and define their profiles, it was decided to conduct an exploratory survey using a structured questionnaire made up of 21 questions, requiring both multiple answers, in some cases with several variables, and single answers. The questionnaire was divided into five sections: production of GL documents; dissemination; bibliographic treatment; production of software tools and reception of documents produced by other organisations. In the questionnaire respondents' personal data was also collected. In this way it was possible to classify the organisations according to the variables identified in the conceptual scheme.

3. Main results of the survey

Data obtained from questionnaires regarding GL producers is analysed in this paragraph, as are the most significant replies from the five sections of the questionnaire (dissemination methods, management of catalogues, production of information tools and frequency with which documents produced by other Organisations are received). Refer to the next paragraph for a more detailed analysis of dissemination modalities. Questionnaire results are analysed with reference to two criteria, reflecting the aims of the survey as a whole: a) to gather information in order to draw up an organisational model from which the future database will be developed, and b) to verify whether there already exists a flow of information between GL producers concerning the exchange of documents.

54.6% of the 898 questionnaires sent out were duly compiled. Most responses came from organisations that perform research activities (66.4%). Among these, the highest response percentages were for peripheral ISPEL Departments (80.6%) and other Research Bodies (75.5%). Significant responses were received from LHUs (50.3% for Internal OSH Responsibles and 57.3% for External OSH Services), which in the past have not been studied in depth and have generally been *unaware* producers of GL, as demonstrated by the numerous contacts made when administering the questionnaire.

Table 1 - Comparison between the number of questionnaires sent, % replies received, % GL producers and % distribution of GL producers by Organisation and function

	No. sent	Responses (%)	GL Producers (%)	Distribution of GL producers (%)
Research				
Universities	168	60.7	56.9	17.0
CNR	41	65.9	48.1	3.8
ISPESL	36	80.6	65.5	5.6
Other Research Bodies	53	75.5	55.0	6.4
<i>Total Research</i>	298	66.4	56.6	32.8
Monitoring				
Regional bodies	31	32.3	70.0	2.0
Internal OSH Responsibles	197	50.3	82.8	24.0
External OSH Services	225	57.3	73.6	27.7
<i>Total Monitoring</i>	454	52.4	77.3	53.7
Consulting				
Trade Unions/Employer's Ass.	108	36.1	82.1	9.4
Professional Association.	39	41.0	87.5	4.1
<i>Total Consulting</i>	147	37.4	83.6	13.5
TOTAL	898	54.6	69.7	100.0

69.7% of organisations that replied to the questionnaire were GL producers. This confirms the presence of a large number of GL producers in the field of health and safety. The highest producer percentages were for consulting organisations (83.6%) and monitoring organisations (77.3%), which also constitutes the largest single group in our study (53.7%). It is interesting to note that categories giving the fewest questionnaire replies contain the highest percentage of GL producers (32.3% for responses from Regional Bodies, 70% of which produce GL; 36.1% for responses from Trade unions and Employers' Association, 82.1% of which produce GL). It may be supposed that in some organisations the propensity to reply is determined by the very fact of being GL producers. In other cases a number of factors may determine the propensity to participate in surveys, for example being accustomed to compiling questionnaires or the ease in identifying the person best suited to supplying the requested information.

3.1 Types of GL Documents

An analysis of GL produced by ISPESL led us to include additional document types in the questionnaire alongside *traditional* GL e.g.: Guidelines, Safety procedures and Technical evaluations. The inclusion of additional GL types derives firstly from the need to use the same terms adopted by producers, who in this case are not particularly familiar with problems associated with GL use. Indeed, whenever specific sectors are analysed [11], further types are often added to the already long list of GL document types.

Figure 3 shows the distribution of GL producers by document type and production frequency. Firstly it shows that production involves types of document closely related to the field. Most organisations produce Course hand-outs (57%), Technical evaluations (51.1%) and Technical reports (50.8%), while Bibliographies (15.2%), Newsletters (12.3%), Journals (10.8%) and

Preprints (5.3%) are produced by fewer organisations. Data regarding Preprints indicates that a very small percentage of organisations produce with a view to possible future publication in a conventional form. This confirms, albeit indirectly, that in the health and safety sector GL is the most common source of information. An intermediate value was obtained for the production of Theses (20.8%).

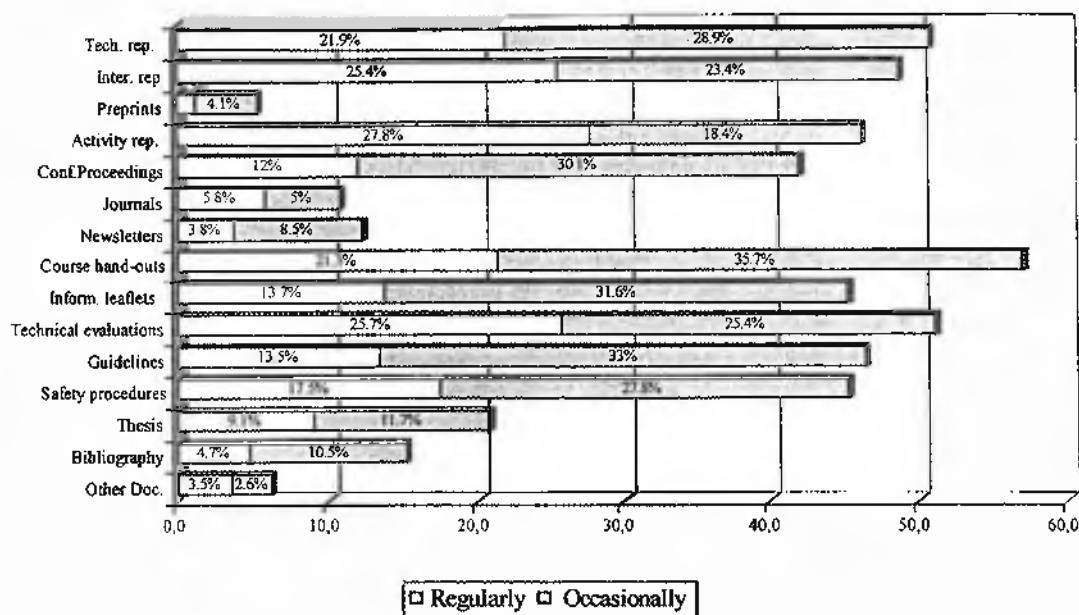


Fig. 3 - Distribution of GL producers by document type and production frequency (%)

Data on the frequency of production also show that a significant percentage of organisations produce some types occasionally, e.g. Conference proceedings (30.1%), Guidelines (33%), Safety procedures (27.8%). This data thus points to specific document types and provides indications on the various activities that play leading roles in this sector: training in the form of Course hand-outs and risk assessment in the form of Technical evaluations, Technical reports, Safety procedures and Guidelines.

3.2 Dissemination of GL

One section of the questionnaire had the specific aim of determining whether organisations were willing to distribute information and identifying the ways in which this is done. (see paragraph 4 for a more detailed look at this section).

The first question asked respondents whether the GL documents produced were distributed mainly to internal users or to users outside the organisation. It emerged that only 7% of respondents do not distribute their own documents, and the breakdown by organisation (Fig. 4) shows that the higher percentages are for University institutes (20.7%) and Regional Bodies (14.3%).

Data on dissemination methods confirm the subdivision by function assumed in the conceptual scheme shown in figure 1. Among the organisations performing consulting activities, 64.3% of Professional associations and 59.4% of Trade unions and Employers' Associations distribute GL documents either to external users only or to both internal and external users. As regards monitoring organisations, there was a further breakdown in distribution tendencies owing to

the specific tasks of monitoring internal safety or the safety of external structures: 57.1% of Regional Bodies and 53.7% of External OSH Services distribute GL both internally and externally, while 81.7% of Internal OSH Responsibles usually distribute GL documents within their own structure in compliance with norms obliging them to disseminate information to workers in the workplace. Among the organisations performing research activities, the role played by ISPESL as a specialised Institute in the sector is confirmed by its internal and

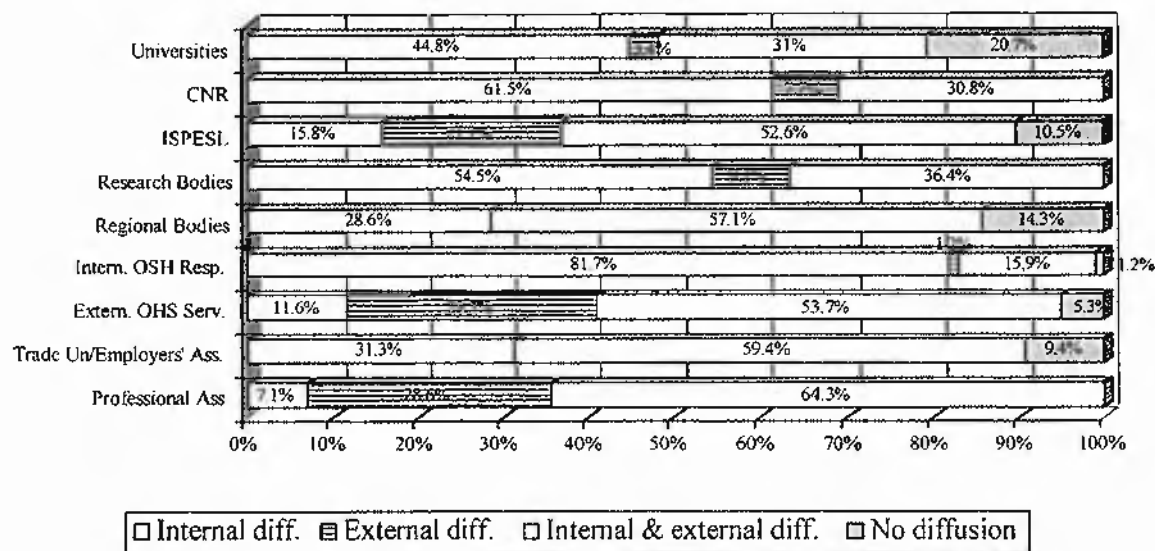


Fig. 4 - % of producers distributing GL

external distribution activities (52.6%), while for University and CNR institutes, and other Research Bodies dissemination is chiefly internal. We assume that organisations which carry out research may prefer other dissemination channels, but this should be checked through a specific analysis of conventional literature published in the field.

3.3 Management of GL catalogues

The respondents were asked to indicate whether the GL they produced was catalogued, and if so, which bibliographic elements were taken into consideration for document treatment. This section of the questionnaire also asked whether catalogues, databases or lists of publications bringing together GL documents had been created. This information, in addition to indicating a certain attention to the management of documentation, gave important indicators regarding the organisation of the future database. If indeed the management of GL were stable, it would be easier to acquire documentation and draw up a shared cataloguing system.

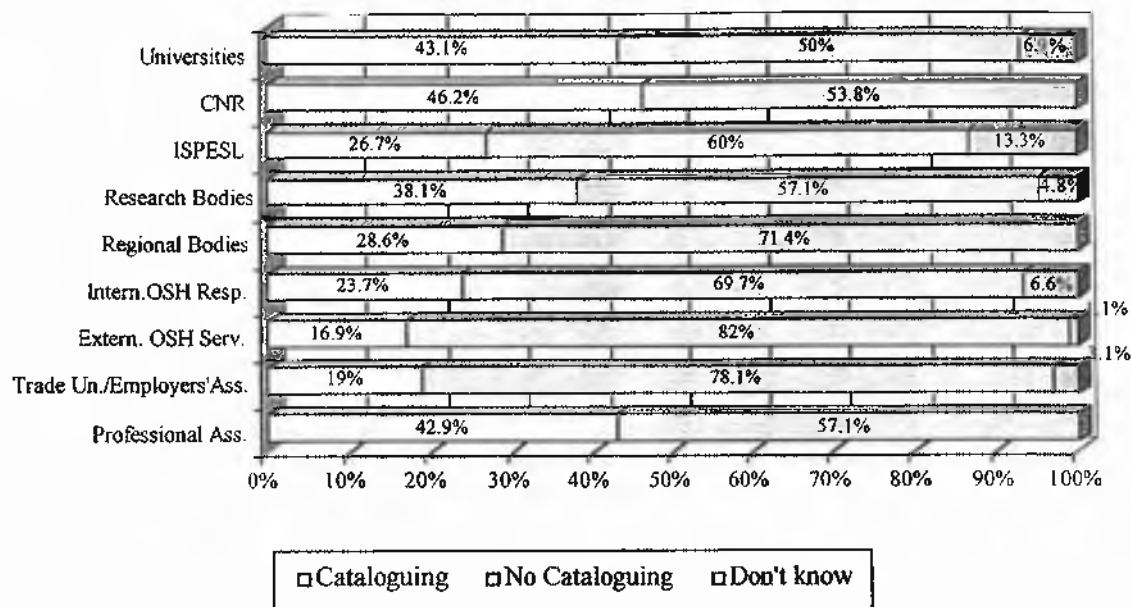


Fig. 5 Management of catalogues broken down by organisation (%)

Only 28.2% of producers actually catalogue GL. Distribution by organisation (Fig. 5) highlights some differences. The highest percentage for organisations cataloguing GL was for University institutes (43.1%), CNR (46.2%), Professional associations (42.9%) and Research Bodies (38.1%). The lowest percentages were for External OSH Services (16.9%) and Trade unions and Employers' Associations (19%).

3.4 Production of "grey" information tools

The fourth section of the questionnaire aimed to verify the existence of information set out in databases, or as hypermedia documents or simulation systems. In this section the respondents were also asked to indicate the media used and topics dealt with. These information tools can also be considered as *grey documentation* or, as it is often called, *grey information* [12] since it is not usually circulated outside the organisation that produces it. These tools are a move forward from *traditional* GL and, if they are produced in the health and safety sector, could be acquired or at least reported in ISPESL's future database.

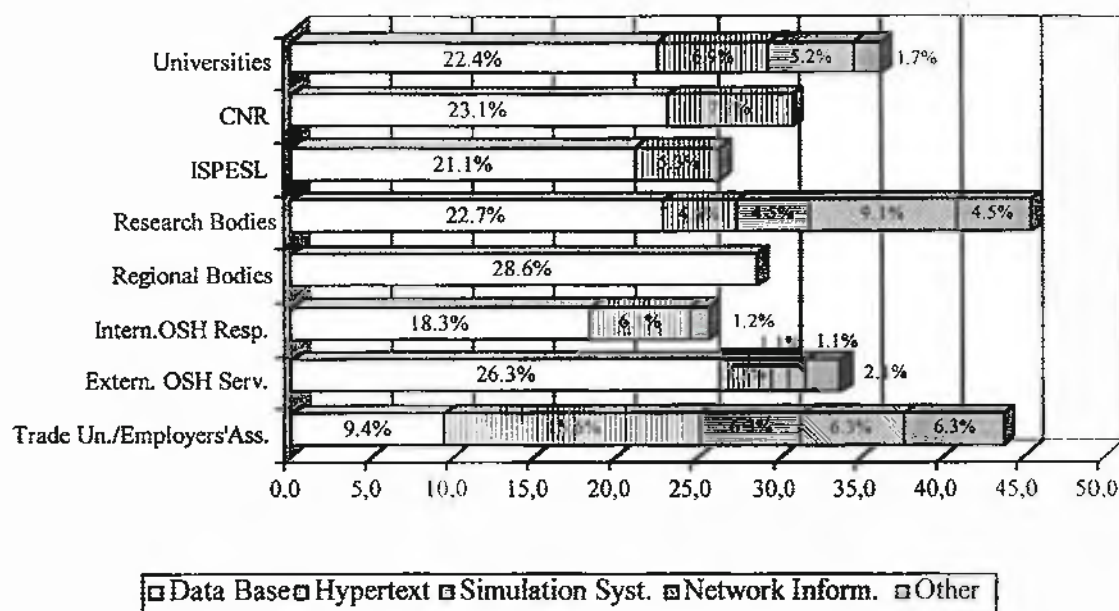


Fig. 6 -Production of information tools broken down by organisation (%)

It emerged that 101 of the 342 GL producers (29.5%) had produced "grey information tools" over the past three years. Most of these organisations produce databases, although percentages were rather low (Fig. 6). The highest values were for Regional Bodies (28.6%) and External OSH Services (26.3%). Trade unions and Employers' Associations (15.6%) specialised more in the production of hypertexts. It should also be noted that most producers of these tools use them to gather statistics and legislative information and that magnetic storage media was usually used. The chart does not show Professional associations since none of these produces any type of information tool.

If these results are compared with data on the format used for produced documents, (58.7% of the respondents produces exclusively paper documentation) with the percentage of organisations that already have a Web page (6%) or that intend to have one (30.4%), we can say that GL producers in the sector do not make full use of the potential of new technologies.

3.5 Reception of GL produced by other Institutions

The final section of the questionnaire concerned the existence of the final link in the information chain, namely the frequency with which producer organisations receive GL documents from other institutions. GL producers are in fact often users of this kind of document, and if they are included in an information flow, they may benefit from an exchange of documents with other organisations.

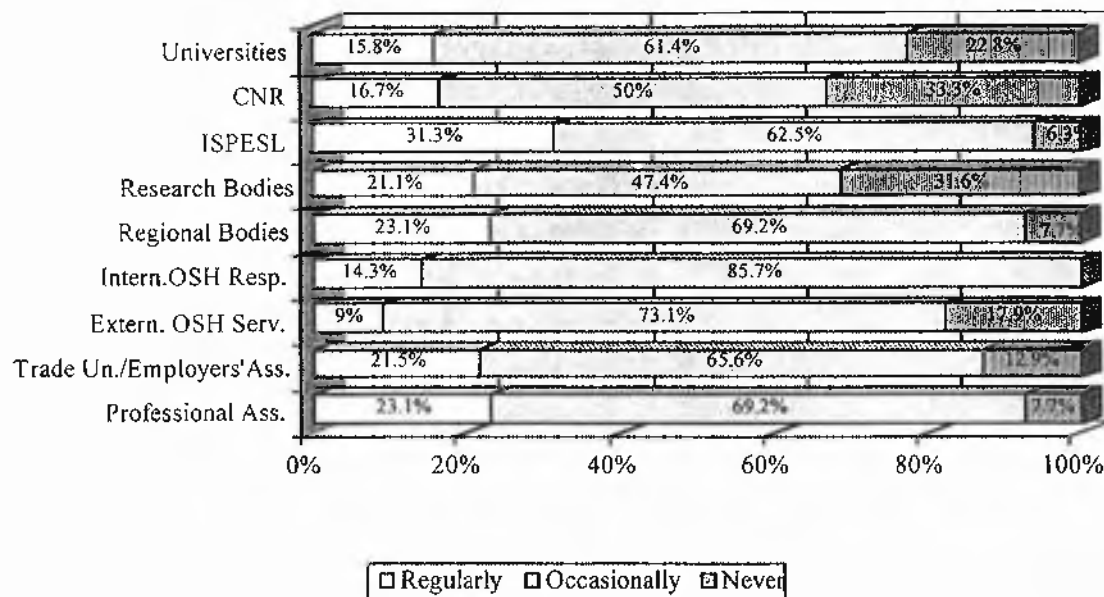


Fig. 7 - GL reception frequency broken down by organisation (%)

Most organisations receive GL from other institutions only occasionally (Fig. 7), while organisations regularly receiving GL are Trade unions and Employers' associations (32.3%) and ISPESL Departments (31.3%). Most organisations performing research activities reported that they had never received GL from other institutions (33.3% of CNR Institutes, 31.6% of Research Bodies and 22.8% of University institutes).

It should be noted that for the final question asking to indicate the names of institutions from which documents are received, many GL producers indicated ISPESL. This bears out ISPESL's role in the health and safety sector and creates the conditions for the success of the future database.

Summing up, for a large number of GL producers in the field of health and safety the data analysed up to this point reveals various profiles, sometimes relating to the organisation's institutional role and sometimes to the more general characteristics of GL. Data on GL production confirms that special types of documents are produced having sector-related information contents. Data on the dissemination of documents produced reflects, for example, the functions performed by these organisations, leading to a greater or lesser propensity to disseminate information internally or externally. Data on the non-management of documentation produced confirm the characteristics common to all GL, for which the absence of bibliographic controls has always been a major obstacle for retrieval and dissemination activities. Data on grey information tools, too, and the predominantly paper-based production of documents are proof of the partial use made of new technologies on the part of GL producers in the sector. These results will influence the organisational model of the future database

4. Dissemination of documents

The section of the questionnaire concerning the dissemination of documents touches upon one of the crucial aspects of GL. The difficulties encountered in obtaining non-conventional literature is indeed due to the non- or limited activation of communication channels and to the under-estimating of potential users. This may be caused by a number of factors, some of which are examined in the questionnaire, such as the inadequacy or absence of internal services to

support this type of activity (library, documentation centre, editorial office or staff assigned to this task), a lack of awareness about the value of information in general and about the content of GL documents in particular, as well as a lack of familiarity in using external information sources and exchanging information, albeit informally.

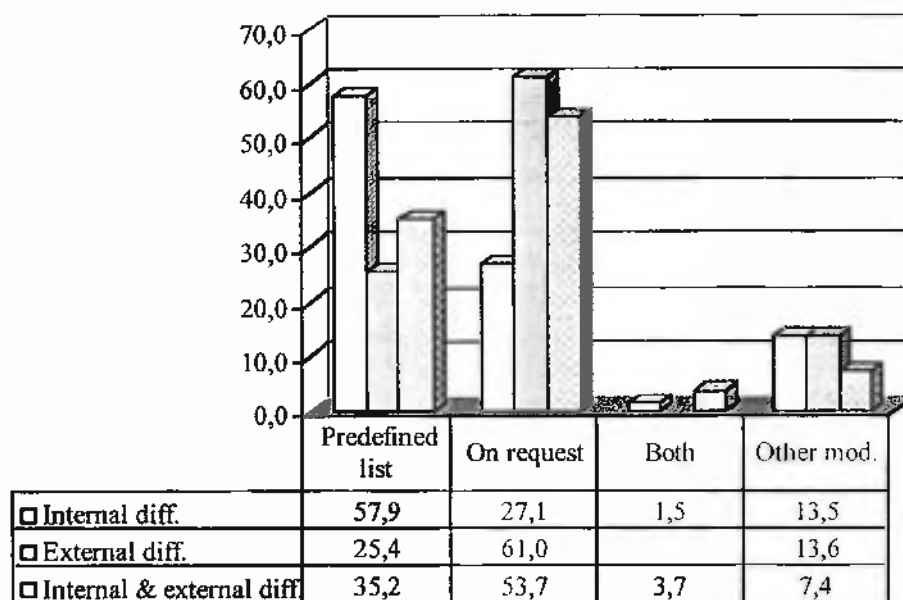


Fig. 8 - GL dissemination modalities (%)

The respondents were asked to indicate whether they distributed their documents through a predefined list or on request, or using both modalities (Fig. 8). This data is compared with results on internal and external diffusion (cfr. par. 3.2).

The use of a predefined list assumes that producers have identified a group of *habitual users*. The data gathered shows that such a list is used by most organisations (57.9%), especially for internal distribution, which, as already mentioned, is of particular significance for the dissemination of information on health and safety in the workplace. The exclusively external distribution, which represent the "expert to expert" dissemination, occurs for the 61% of organisations on request. This means that there are no fixed means of dissemination, nor are there established channels for the exchange of documents.

4.1 Information flows

GL producers were asked to indicate the institutions to which their documents are distributed (Central State Administration, Regional and Local Administrations, Universities, Research centres, Trade unions, Professional associations, Private companies, own members, own employees). It was thus possible to verify whether there is already a flow of information among those GL producers chosen for the survey, and whether such a flow gives priority to particular types of institutions.

Table 2 shows an input-output grid relating organisations that send out their documents

(output documents) to those receiving them (input documents).

The institutions indicated by producers as the receivers of their documents were grouped according to their main role using criteria similar to those adopted for GL producers. Processed data did not therefore take into account replies indicating central administrations and private companies, which were not included in the survey, while organisations indicating that they distributed GL to their own members and/or employees were classed as belonging to internal flows, that is to information flows directed towards the same type of organisation.

Table 2 - Information flow among GL producers broken down by function (%)

Sending Institutions	Institutions receiving GL		
	Research	Monitoring	Consulting
Research	56.1	19.3	24.6
Monitoring	4.8	35.7	59.5
Consulting	5.7	10.5	83.8

The information flow given in table 2 shows that documents are exchanged chiefly within organisations covering similar roles, and this is particularly evident for organisations performing consulting activities (83.8%). Exceptions were organisations performing monitoring activities, which are those showing the greatest flows of information (59.5%) towards organisations performing consulting activities. This may be explained by the need to communicate the results of their monitoring and supervisory activities to consulting organisations. Organisations performing research activity receive little input from organisations with different roles, while their output is higher (24.6% and 19.3%). It should be noted however that these figures are affected by answers given by peripheral ISPESL Departments that operate closely with local contexts and, as they belong to a national organisation expressly dedicated to the sector, perform liaison activities among the various health and safety services.

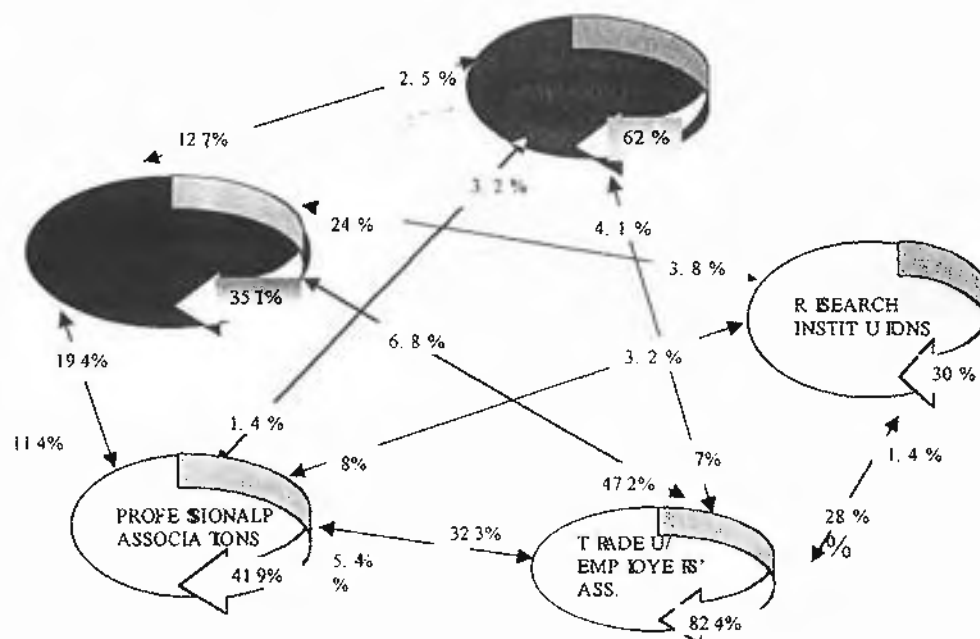


Fig. 9 - Information flow of GL producers broken down by organisation

The information flows broken down by organisation (Fig. 9) reinforces the idea of an exchange of documents limited to organisations of the same type. This is especially evident for Universities (62%) and Trade unions and Employers' Associations (82.4%). The State's Peripheral Administrations, including Regional Bodies, as well as Internal OSH Responsibles and External OSH Services, communicate not only internally (35.1%) but above all with Trade unions and Employers' Association (47.2%).

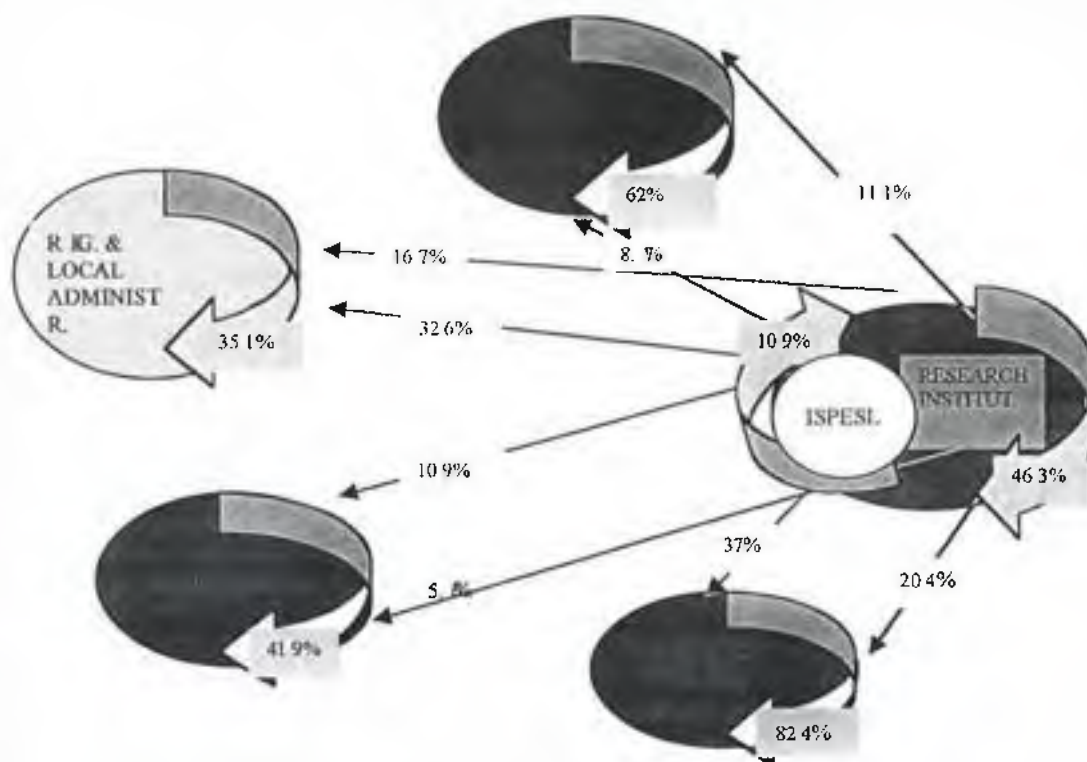


Fig. 10 - ISPEL document output compared with other organisations

Figures on Research Institutes, comprising CNR Institutes, other Research Bodies and ISPEL Departments, are again influenced by answers given by ISPEL. If one analyses individual data (Fig. 10), it appears that the information flows of ISPEL Departments are chiefly directed towards the Peripheral Administration (32.6%) and towards Trade unions and Employers' Associations (37%), while of much lesser importance are flows to Universities (8.7%) and other Research Institutes (10.9%). On the other hand the information flows between Research Institutes not comprising ISPEL Departments again highlight the fact that the exchange of information occurs chiefly among organisations of the same type: flows between Research Institutes were 16.3% higher in this case, going from 30% to 46.3% (cfr. Fig. 9). Excepting the exchange of information between Research Institutes and Universities, flows to other organisations were considerably lower.

Support structures for document editing, cataloguing and dissemination activities

GL producers were asked whether their organisations possessed structures set up to perform GL document editing, cataloguing and dissemination activities, and if so, to indicate one of those listed (library, documentation centre, archive department, editorial office, press office, publications office, public relations office). 57% of Organisations replied positively to the first question, but in the follow-up question added an unforeseen item: "document producer", turning around the meaning of the question.

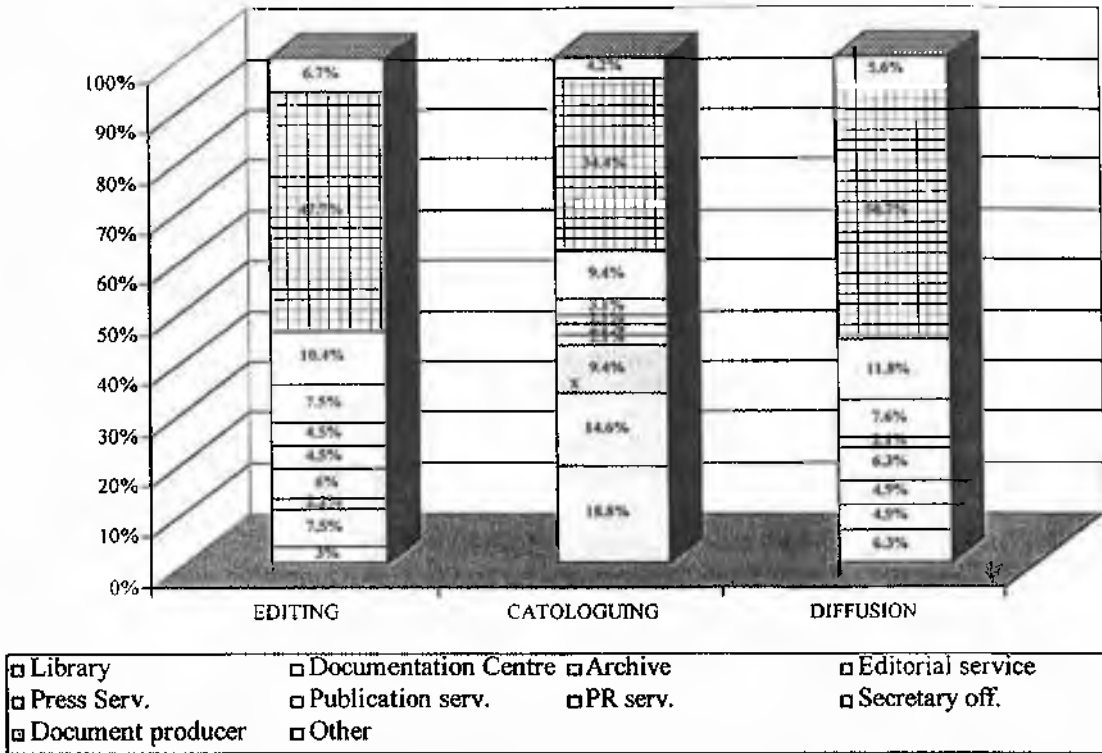


Fig. 11 – Presence of departments performing GLediting, cataloguing and dissemination activities (%)

In most organisations it is the author him/herself (the "document producer") who drafts the document (47.7%) and distributes it (50.7%). Cataloguing activity is performed as follows: 34.8% by the document producer, 18.8% by the library, 14.6% by the documentation centre. This fact will render the acquisition of documents for the future database more complex.

On the whole, data regarding dissemination highlights the existence of well-identified users within organisations, receiving information on prevention and safety measures. On the other hand, external dissemination is prompted more by users requesting information than by established dissemination practices. Despite this fact, most producers indicated that they send their GL documents to other institutions, and resulting information flows are directed largely towards organisations of the same type. Only ISPESL Departments and Regional and Peripheral Administrations have opened up information channels to organisations with different roles. Finally, very few organisations make use of support services for editing, cataloguing and dissemination activities, and most of these activities are performed directly by the document's author. This result is perhaps not so surprising for GL, in which the author plays an active role in each phase of the document's life cycle, becoming the main distributor of produced documents, especially in sectors making use of online communications.

5. Conclusions

The survey confirms the hypothesis that a large number of organisations in the field of occupational health and safety produce GL, and that the development of a database containing such documents may be a useful tool for improving communication among different organisations, increasing information dissemination modalities and accordingly enhancing

information channels between organisations with different roles.

The survey looked at a range of different contexts in terms of the institutional mission, organisational set-up and role played by each organisation. The resulting profile of the GL producer shows specific characteristics especially for the production of field-related document types. Features shared by most organisations include the modalities used to manage and disseminate GL production. Only a small number of organisations have arranged for the bibliographic management of GL production, and dissemination is generally directed towards internal users. External distribution appears to be unplanned, information flows being directed towards organisations performing similar role. Finally, a limited number of organisations have services supporting document editing, cataloguing and dissemination activities.

These results confirm that once again the difficulty in retrieving GL and its limited dissemination derive from the low priority which producer organisations give to the exchange of information, and of the degree of awareness of its information value.

This fact has a direct influence on the design and development of the future database, since it will be necessary to choose an organisational model for the acquisition and bibliographic management of documents that can fit in with the needs of both producers and database managers. As indeed the success of the project depends on the degree of cooperation among producer organisations, which will moreover constitute a sizeable portion of future database users, a modular, flexible system will have to be developed. In particular it will be necessary to provide simple document-cataloguing methods, whenever possible, while access to the database should not be based solely on network distribution as most organisations do not yet possess such computer-based instruments.

The project is now in the data modelling and software selection phase. There are also plans to illustrate the results of the survey to those Organisations that replied to the questionnaire to raise awareness and involvement in order to build an effective communication and information network.

References

- [1] Decreto Legislativo 19/9/94, N. 626. Attuazione delle direttive 89/391/CEE, 89/654/CEE, 89/655/CEE, 89/656/CEE, 90/269/CEE, 90/270/CEE, 90/394/CEE e 90/679/CEE riguardanti il miglioramento della sicurezza e della salute dei lavoratori sul luogo di lavoro. *Gazzetta Ufficiale della Repubblica Italiana*, n. 141, 12 Nov. 1994.
- [2] Council Directive of 30 Nov. 1989 concerning the minimum safety and health requirements for the workplace [CEC]. *Official Journal of the European Communities*, 30 Dec. 1989, vol. 32, no. L393, p.1-12.
- [3] Castriotta, M. La documentazione di tipo sanitario a misura del paziente e del lavoratore. In: *Atti del 5. Convegno Nazionale AIDA "Documentazione: professione trasversale"* a cura di C. Basili, Roma, 1998, p. 163-170.
- [4] Holzer, C.L., Hirsch, R.B. Managingn workplace exposure information. *American Industrial Hygiene Association Journal*, 54 (1), 1993, p.15-21.
- [5] Auger, C.P. Information sources in grey literature. 3rd ed., London: Browker-Saur, 1993.
- [6] Farace, D.J. Rise of the Phoenix: A review of new forms and exploitations of Grey Literature. *Publishing Research Quarterly*, 13 (2), 1997.
- [7] Hart, G.W. Aerospace and other grey literature from Nato. In: First International Conference on Grey Literature, Amsterdam, 13-15 dec. 1993. Amsterdam, Organising Bureau, 1994, pp. 18-32.
- [8] Rozkuszka, D.W. The art and acquisition of foreign official publications. In: Official publications of Western Europe, vol. 1 Edited by Eve Hohansson. London: Mansell Publ.Lim., 1984
- [9] Sprehe, T.J. Government information: from inaccessibility to your desktop and back again. *Journal of the American Society for Information Science*, 50 (4) 1999, p. 340-345.
- [10] Kreitz, P.A., Addis, L., Galic, H. Johnson, J. The virtual library in action: collaborative international control of High-Energy Physics pre-prints. In: 2nd. Conference on Grey Literature, Washington, 2-3 Nov. 1995. Amsterdam, Organising Bureau, 1996, p. 33-41.
- [11] D.J. Farace. Glossary, a vocabulary of terms in GL. Grey Literature Network Service. Amsterdam: GreyNet, 1999, 60 p.
- [12] Carroll, B.C., Cotter, G.A. A new generation of Grey Literature: the impact of advanced information technologies. In: First International Conference on Grey Literature, Amsterdam, 13-15 dec. 1993. Amsterdam, Organising Bureau, 1994, pp. 5-17.

Preserving the Pyramid of STI Using Buckets

Michael L. Nelson and Kurt Maly

ABSTRACT

The product of research projects is information. Through the life cycle of a project, information comes from many sources and takes many forms. Traditionally, this body of information is summarized in a formal publication, typically a journal article. While formal publications enjoy the benefits of peer review and technical editing, they are also often compromises in media format and length. As such, we consider a formal publication to represent an abstract to a larger body of work: a pyramid of scientific and technical information (STI). While this abstract may be sufficient for some applications, an in-depth use or analysis is likely to require the supporting layers from the pyramid.

We have developed buckets to preserve this pyramid of STI.¹ Buckets provide an archive- and protocol-independent container construct in which all related information objects can be logically grouped together, archived, and manipulated as a single object.² Furthermore, buckets are active archival objects and can communicate with each other, people, or arbitrary network services. Buckets are an implementation of the Smart Object, Dumb Archive (SODA) DL model. In SODA, data objects are more important than the archives that hold them. Much of the functionality traditionally associated with archives is pushed down into the objects, such as enforcing terms and conditions, negotiating display, and content maintenance. In this paper, we discuss the motivation, design, and implication of bucket use in DLs with respect to grey literature.

1
Digital
Library

INTRODUCTION

Research projects produce information in a variety of formats. A traditional formal publication, such as a journal article, is generally supported by a large quantity of software, datasets, images, video, informal notes, presentations and other documents. Collectively, we call this set of scientific and technical information (STI) the “Pyramid of STI”.

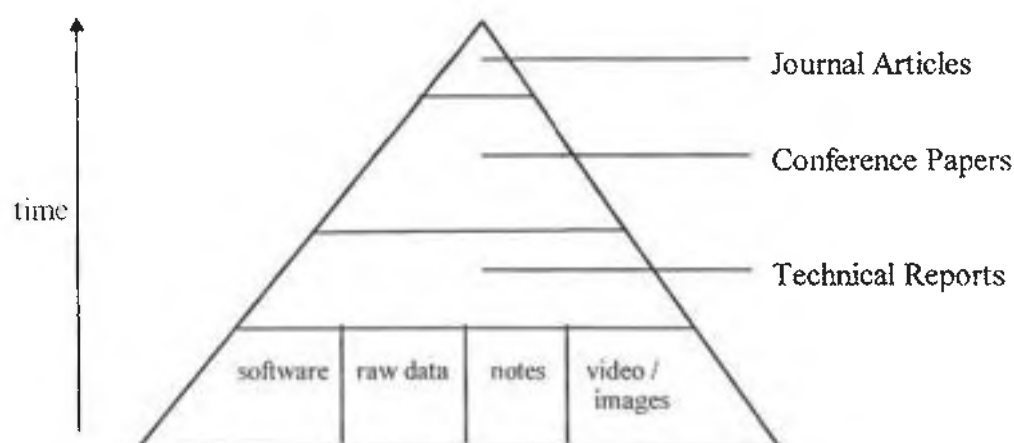


Figure 1: Formal Publications Rest on a Pyramid of STI

Although the information in this pyramid was created together and subtle relationships between its components can exist, different semantic instantiations are generally segregated along currently obsolete media boundaries. Reports are placed in report archives, software might go into a software archive, but most of the data, supporting materials and other grey information are likely to be kept in informal personal archives or discarded altogether. Our experience with NASA digital libraries (DLs) is users wish to have access to the supporting materials, data and software used in the preparation of formal literature (Sobieszczanski-Sobieski, 1994) ... even though there currently is no well established publication vector for much of this information. ✓

We feel many DL projects focus on simply automating the formal publication process, while providing little attention to the lower tiers of the pyramid of STI. Similarly, we feel that creating “separate but equal” DLs implemented congruently with obsolete media boundaries pushes the burden of STI (re)integration to the user.

To address these concerns, we have created buckets: aggregative, intelligent agents tailored for publishing in digital libraries. Buckets can be thought of DL-specific Digital Objects as described in (Kahn and Wilensky, 1995). Buckets provide a container mechanism to capture and preserve the pyramid of STI. To enhance their long-term survivability, buckets are completely self-contained. [All the logic to manage, and protect their content is embedded in the bucket itself, not in a separate server.] Buckets can also be thought of as "archivelets". Communication with the bucket occurs through messages utilizing the bucket application programming interface (API). [By promoting buckets to first class network citizens, we are stressing that content is more important than the search engine or DL protocol used to access it.] By imbuing the content with additional functionality, we increase the long term usefulness and survivability of the STI.

Grey Literature at NASA

NASA communicates its research findings through the traditional open literature process as well as its own multi-tiered, self-published report series (Pinelli, 1990). The NASA report series offers a number of advantages to authors: no page restrictions, potential for restricting dissemination, possibility of color graphics, and occasionally the inclusion of a CD-ROM of data, images or software. However, the latter two are rarer than most authors would like because they are expensive to create, and their distribution is more expensive still. The NASA reports are often ingested in systems that can handle only paper hard copy or possibly just microfiche -- leaving few options for propagation of additional media formats such as CD-ROMs.

An even more compelling case for capturing grey literature at NASA is that the formal publications (NASA's report series or open literature) represent a decreasing percentage of the total amount of STI created and used by NASA and its customers. [Due to the increasingly proprietary nature of NASA's work, as well as increasing time constraints on fewer staff members, many research projects are no longer resulting in a formal publication.] Instead, the projects remain as a collection of briefings, data, and other forms of grey literature -- often with proprietary access restrictions. While neglecting the formal publications achieves the short term goal of increased project turn around time, the inability to capture and preserve any

of the resultant STI creates a gap in the corporate memory. There is no well-defined, large scale publishing outlet for a majority of the STI created at NASA.

Shortcomings of Current Digital Libraries

We recently surveyed a number of digital libraries from multiple disciplines and found that most focused only on formal publications (Esler & Nelson, 1998). Even DLs that focus on grey literature (i.e. technical reports, pre-prints, working papers) still only focus on the hard-copy representation. Software, images, video and other material, if they are preserved at all, are often sent to DLs that serve only those respective media formats. Such examples include NIX, an image DL for NASA (von Ofenheim, et al., 1998); Netlib, a software DL for the high performance computing community (Browne, et al., 1995); and Alexandria, a DL for geospatial data (Smith, 1996). While these and other like DLs can provide custom interfaces for interacting with non-textual STI, they have as a side effect an artificial segregation along media formats.

We are aware of no other current DL that allows for archiving and serving a complex set of STI. Consider a research project that produced raw data, software to reduce the data, reduced data, shift notes, image or video representation of data, presentations, informal reports, and a journal article that summarized the project's findings. Currently, the journal article has the best chance of appearing in a DL. Possibly the software will make it to a software DL, and depending on the discipline the data might make it to a data archive, but neither is guaranteed. The "base" of the pyramid of STI is at worst, effectively thrown away, and at best is splintered and sent to separate DLs. If researchers read the journal article and wish to extend the data analysis differently than presented, they can either search in other DLs for the software and datasets, or they can contact the authors in the hopes of obtaining the software and datasets through collegial distribution. The former assumes the various non-report STI has been placed in DLs somewhere, while the latter assumes the authors can still be reached.

BUCKETS

- ¹¹ Buckets are object-oriented container constructs in which logically grouped items can be collected, stored, and transported as a single unit. ¹²Buckets are completely self-contained: they

have their content physically resident, the intelligence to serve, manage and enforce terms and conditions for accessing the content. Their self-contained nature allows them to function independently of, and in cooperation with, any DL system or protocol. It also allows them to be completely mobile - a bucket can "move around" and still retain its functionality. Buckets also maintain their own logs of actions performed on them, so they can retain a history that is independent of the servers used to access them.

As far as what goes into a bucket, buckets provide mechanism -- not policy. There are no pre-defined concepts of what information types should go into a bucket, and buckets make no assumptions about their content. The authors and publishing organizations control what constitutes a bucket; if there is reason why a lower strata in the pyramid should not be preserved, there is no requirement that it be included. Similarly, buckets are not unsuitable for white literature; they simply offer the capability (should the authors and publishers choose) to extend beyond the white into the grey.

* To the casual observer, a bucket appears as a regular web page. However, messages are sent to the bucket using the hypertext transfer protocol (http). If no message is sent, by default the bucket builds an HTML presentation page of its contents:

<http://dlib.cs.odu.edu/test-bucket3/>

However, other messages are possible. If a web robot wanted to gather the structured metadata, it could issue this bucket message:

<http://dlib.cs.odu.edu/test-bucket3/?method=metadata>

To learn what methods this bucket supports, it could issue:

http://dlib.cs.odu.edu/test-bucket3/?method=list_methods

Many other methods are defined, including methods listing the bucket source code, updating the bucket source code, changing who are privileged principals to the bucket, and other various functions. See (Nelson, et al., 1999) for a further discussion of bucket methods and implementation.

SODA: SMART OBJECTS, DUMB ARCHIVES

An observation from our experiences with current NASA DLs is that a surprising number of people do not find the publications via the respective DLs. Since the full contents of the NASA DLs are browsable, both the abstract lists and the reports are indexed by web crawlers, spiders and the like. Users are formulating complex queries to services such as Yahoo, Altavista, Lycos, Infoseek, etc. to find NASA STI. We presume this is indicative of the resource discovery problem: people start at these portals because they do not know all the various DLs themselves; and the meta-searching problem: they are trusting these services to search many sources, not just the holdings of a single DL.

Although we believe we have built attractive and useful interfaces for the NASA DLs, our main concern is that people have access to NASA's holdings and not that they use a given DL interface. It is desirable that NASA publications are indexed by many services. Since there can be any number of paths to the information object, the information object must be a first class network citizen, handling presentation, terms and conditions, and not depending on archive functionality.

In the SODA model, we have separated DLs into three separate layers: 1) the Digital Library Services (DLS) layer that provides user interfaces such as searching and browsing; 2) the archive layer that manages collections of objects; and 3) the objects themselves. Separating the functionality of the archive from that of the DLS allows for greater interoperability and federation of DLs. The archive's purpose is to provide DLs the location of buckets (the DLs can poll the buckets themselves for their metadata), and the DLs build their own indexes. And if a bucket does not "want" to share its metadata (or contents) with certain DLs or users, its terms and conditions will prevent this from occurring. For example, we expect the NASA digital publishing model to begin with technical publications, after passing through their respective internal quality control, to be placed in a NASA archive. The NASA DL (which is the set of the NASA buckets, the NASA archive(s), the NASA DLS, and the user communities at each level) would poll this archive to learn the location of buckets published within the last week. The NASA DL could then contact those buckets, requesting their metadata. Other DLs could index NASA holdings in a similar way: polling the NASA archive and contacting the appropriate buckets. The buckets would still be stored at NASA.

but they could be indexed by any number of DLs, each with the possibility for novel and unique methods for searching or browsing. Or perhaps the DL collects all the metadata, then performs additional filtering to determine applicability for inclusion into their DL. In addition to an archive's holdings being represented in many DLs, a DL could contain the holdings of many archives. If we view all digitally available publications as a universal corpus, then this corpus could be represented in N archives and M DLs, with each DL customized in function and holdings to the needs of its user base. The SODA model for DLs is discussed in detail in (Maly, et al., 1999a).

FUTURE WORK

There are several areas we continue to explore and develop, the first of which is tools for buckets. The long-term success of buckets will depend on the quality of the tools to create and manage buckets. We have Publishing Tool, that allows users to transparently create and populate buckets. We have a Management Tool, which is a simple workflow mechanism to review buckets submitted for publication, approve or reject them, and move them into designated areas when approved. Finally, we have an Administration Tool that allows for long term maintenance of buckets, performing such operations as large scale source code and principal updates. These tools are undergoing frequent revisions as we gain user feedback.

Secondly, we are developing the concept of discipline specific buckets that can exploit knowledge about their contents. These bucket templates could have specialized display capabilities, specialized methods for data interaction, custom storage options, etc. These would be in contrast to the ontology neutral buckets presented here. The first of these, a bucket template for undergraduate education, is discussed in (Maly, et al., 1999b).

The third area we are actively developing is making buckets intelligent agents. The buckets are already carrying the intelligence to manage their content -- why not make them even more intelligent? We are designing a Bucket Communication Space (BCS) that will allow buckets to communicate with other buckets, people, and arbitrary third party network resources. The BCS will allow for functions such as messaging, format conversion, and matching. The latter is especially interesting: we will allow buckets to find similar buckets off-line. Buckets will register their profiles with the BCS, and the BCS will inform buckets of their potential matches

based on a similarity index. Buckets can then either automatically establish linkages, or contact their owners for verification.

Additional functions afforded by intelligent buckets includes a redefinition of software reuse. If general software resides in a bucket in a DL, users can either download and integrate it with their applications, or use the software while it is resident in DL. Many DL projects have agent technology designed to assist the user in searching, browsing or other functionality. However, we are unaware of other projects that attempts to make the archived object itself intelligent. The full implications of making archival objects intelligent agents has yet to be explored.

CONCLUSIONS

Past media format limitations have defined our view of an archival unit of STI. We claim the formal publications that are the focus of traditional libraries and many DL projects are simply an abstract to a larger body of information. This pyramid of STI is supported by informal documents, software, data, images, videos and other multi-format material. Our DL experiences have lead us to believe that there is high user interest in obtaining access to the lower tiers of the pyramid of STI -- many of which are not archived at all.

To this end, we have developed buckets: object-oriented, intelligent agents for publishing in DLs. Buckets are DL system and protocol neutral; they depend on no particular system and are designed to have minimal impact on any such system. Messages are communicated to buckets via an API that uses http as a transport. To increase their long term survivability, buckets are completely self-contained, carrying all of their contents internally as well as the logic to manage, serve and protect those contents. In addition to the aggregative properties of buckets, we are also working to make them intelligent agents. We are implementing the Bucket Communication Space to allow buckets to communicate with each other and perform tasks such as finding and linking to similar buckets. By imbuing STI objects with aggregation and intelligence, we can now preserve the entire pyramid of STI, as well as introduce new functional capabilities for archived objects.

REFERENCES

- Browne, S.; Dongarra, J.; Grosse, E. and Rowan, T. (1995), "The Netlib Mathematical Software Repository", *D-Lib Magazine*, September 1995. Available:
<http://www.dlib.org/dlib/september95/netlib/09browne.html>
- Esler, S. L. and Nelson, M. L. (1998), "Evolution of Scientific and Technical Information Distribution", *Journal of the American Society of Information Science*, Vol 49 No 1, pp. 82-91.
- Kahn, R. and Wilensky, R. (1995), "A Framework for Distributed Digital Object Services", *cnri.dlib/tn95-01*. Available:
<http://www.cnri.reston.va.us/cstr/arch/k-w.html>
- Maly, K.; Nelson, M. L. and Zubair, M. (1999), "Smart Objects and Dumb Archives: A User-Centric, Layered Digital Library Framework", *D-Lib Magazine*, Vol 5 No 3. Available:
<http://www.dlib.org/dlib/march99/maly/03maly.html>
- Maly, K.; Zubair, M.; Liu, X.; Nelson, M. L., Zeil, S. J. (1999), "Structured Course Objects in a Digital Library", to appear in *Proceedings of the International Symposium on Digital Libraries*, Tsuka Japan.
- Nelson, M. L.; Maly, K.; Zubair, M. and Shen, S. N. T. (1999), "Buckets: Aggregative, Intelligent Agents for Publishing", *Webnet Journal*, Vol 1 No 1, pp. 58-66.
- Pinelli, T. E. (1990), "Introduction to National Aeronautics and Space Administration's Scientific and Technical Information Program", *Government Information Quarterly*, Vol 7 No 2, pp. 123-126.
- Smith, T. (1996), "A Brief Update on the Alexandria Digital Library Project", *D-Lib Magazine*, March 1996. Available:
<http://www.dlib.org/dlib/march96/briefings/smith/03smith.html>

Sobieszczanski-Sobieski, J. (1994), "How to Improve NASA-Developed Computer Programs", NASA CP-10159, pp. 58-61.

von Ofenheim, W. H. C.; Heimrel, N. L.; Binkley, R. L.; Curry, M. A.; Slater, R. T.; Nolan, G. J.; Griswold, T. B.; Kovach, R. D.; Corbin, B. H. and Hewitt, R. W. (1998), "NASA Image eXchange (NIX)", NASA/TM-1998-206295.

Michael L. Nelson
NASA Langley Research Center
MS 158
Hampton, VA USA 23681
+1 757 864 8511
+1 757 864 8342 (fax)
m.l.nelson@larc.nasa.gov
<http://home.larc.nasa.gov/~mln/>

Michael is an electronics engineer at NASA Langley Research Center and a Ph.D. candidate in computer science at Old Dominion University.

Kurt Maly
Department of Computer Science
Old Dominion University
Norfolk, VA USA 23592
+1 757 683 4817
+1 757 683 4900 (fax)
maly@cs.odu.edu
<http://www.cs.odu.edu/~maly/>

Kurt is the Kaufman Professor and Chair of the Computer Science Department of Old Dominion University.

AN ARCHITECTURE FOR GREY LITERATURE IN A R&D CONTEXT

Prof. Keith G Jeffery

Acting Director Information Technology

CLRC-Rutherford Appleton Laboratory

Chilton, Didcot, OXON OX11 0QX UK

(keith.g.jeffery@rl.ac.uk)

Abstract

The importance of Grey Literature is becoming increasingly recognised. For many organisations it encapsulates the knowledge and know-how and thus is a vital business asset. It has similar importance in quality of life aspects - healthcare, environment and culture. Grey Literature in a R&D environment represents the cutting edge of this knowledge and so its management is of utmost importance. Partly based on involvement in defining the datamodels for R&D information interchange across Europe, the author here defines a content metadata datamodel for Grey literature which is more expressive and has more flexibility than any previous proposal and which integrates seamlessly with the CERIF2000 definition which will soon replace the CERIF1991 European Union Recommendation to Member States on Exchange of R&D Information. The content metadata datamodel offers significant advantages over Dublin Core yet can generate Dublin Core if required.

Keywords:

grey literature, metadata, W3C, catalog, digital library, CERIF, ERGO, datamodel, workflow, Dublin Core

INTRODUCTION

R&D Information

It is generally acknowledged that easy access to R&D information provides innovation motivation for industry and commerce leading to wealth creation and employment, and simultaneously provides quality of life advantages especially in healthcare, environmental and cultural aspects. Provision of such access is particularly interesting in a multilingual, multicultural environment such as Europe.

Work on many aspects of R&D information has been going on for many years, and systems have been developed and are in use. Integration of such systems has been an interest of the author for many years. Now that - through the CERIF and ERGO initiatives - integration of information on R&D projects across Europe is well underway, and integration of information for persons and organisations planned together with links to patents, products and publications, it is time to consider how best to address Grey Literature. Some attempts have been made to integrate Grey Literature into conventional library electronic catalogue systems e.g. (BibSys) and a particular catalog format was adopted for SIGLE (Farace, 1990). The system for handling theses at Virginia Tech (Virginia) has defined its own metadata including that for controlling workflow. However, these attempts do not satisfy the requirements identified.

It is becoming increasingly clear that in many organisations the Grey Literature documents as information, the knowledge and know-how of the organisation - assets that are extremely valuable - and therefore which need to be curated carefully and used for trade with other organisations subject to security and access controls. In a R&D environment such knowledge is at the cutting edge of progress and is thus extremely valuable.

The end-user requirement is for the relevant information (relevance, recall), at the right place (wherever worldwide), at the right time (when required), in the appropriate form (optimal presentation, integrated for further use in electronic information / office environments). The end-user requires this information to be available through an easy-to use, tolerant, intelligently-assisting, multilingual, multimedia, multimodal user interface. Briefly, an interface that obeys 'do what I need, not what I say'.

History of the Paper

The author has designed R&D systems for UK Research Councils and currently leads the project integrating UK R&D resources for presentation to common EU standards. He was involved in the definition of CERIF1991. He is currently designing a system for Grey Literature at CLRC-RAL, having experimented with various digital library environments including within the (DELOS) project over many years. The author has also been associated with work on metadata for R&D projects within the (FORSKDOK) project at University of Bergen for some time. Grey literature has been recorded in the (BibSys) System with limited success.

The author represents UK at the European Union CERIF (Common European Research Information Format) Revision Working Group where the topic of Grey Literature has yet to be addressed - although most other aspects of R&D information have already been discussed and standards agreed - largely based on the datamodels proposed by the author. Thus in the second half of 1998 the author decided to work intensively on Grey Literature; because of the ongoing cooperation with the team there led by Jostein Hauge, a Workshop was organised in Bergen in December 1998. An early version of some of the ideas in this paper was presented there.

Structure

The rest of the paper is organised as follows: Section 2 reviews the background state of R&D information concluding with the urgent need for systems for Grey Literature. Section 3 discusses requirements of a Grey Literature System and Section 4 develops an architecture. Section 5 discusses options in implementation and presents a datamodel for Grey Literature Content metadata which is a superset of other content metadata standards and proposals. Current work on other kinds of metadata is outlined. Section 6 concludes.

BACKGROUND : R&D SYSTEMS

The Grey Literature requirement here is viewed as one component of a much larger environment for R&D (Research and Development) information. This implies that a Grey Literature Resource (object) will have relationships with persons, organisational units, projects, patents, products, other publications.

The Customers / Users

The customers / users of a R&D system and their particular requirements include:

- (a) Researchers: find partners, locate techniques, manage projects, produce outputs;
- (b) University Officials: authorise, examine, produce statistics;
- (c) Research Council (R&D Government funding source) officials : evaluate projects and results, produce statistics;
- (d) Commerce & Industry: find ideas, technologies of potential relevance;
- (e) Intermediaries: assist commerce and industry in finding relevant innovative technologies;
- (f) Media: find news articles;
- (g) General Public: general awareness of R&D.

The Products of R&D

The products of R&D are listed below. Items in **bold** are regarded as 'grey literature': there is a need to design to manage all of these kinds but it was decided to concentrate on publications first:

- (a) publications: journal / conference papers, **technical papers, theses, dissertations, reports**
- (b) patents
- (c) products: **prototypes**, fully engineered products
- (d) results: **data and its presentation / visualisation**
- (e) know-how and IPR: **reports, procedure instructions**
- (f) education and training: **documentation, courses**
- (g) publicity: **press releases, product or organisational 'flyers'**

R&D Project

The products are the output from a project, which has a mission, objectives, deliverables, project plan, a duration and resources. The project has a need for access to pre-existing and current information. The project goes through the stages: proposal / project execution / evaluation. Additionally, Researchers need project management / accounting tools but these are not considered here.

Pre-Existing R&D Systems

It is pointless, both for legacy-usage and economic reasons, to re-develop in a R&D environment systems for patents, products, publications; they exist and they work. Of course there are shortcomings and criticisms but in general the systems provide a service e.g library catalog systems. Hopefully they will evolve to continue to meet user requirements, if not replacement will have to be judged on the business case at the time.

However, it should be noted that existing systems are distributed, sometimes unconnected and heterogeneous in the following characteristics:

- (a) purpose and intent;
- (b) structure (relationships);
- (c) content (attributes, language, character set);
- (d) available input / output interfaces;

so it is necessary to provide interoperability interfaces in order for the end-user to have an integrated information landscape. This is achieved using metadata and associated processing.

Pre-existing systems for R&D projects also exist; they, too, are distributed and heterogeneous and the ERGO project utilising the CERIF1991 Recommendation is integrating across this heterogeneity using an enhancement of the the catalog technique (Jeffery, 1992) as used in IDEAS (Jeffery et al, 1989) and the subsequent EXIRPTS projects - thus providing a common datamodel. The CERIF2000 recommendations formalise and improve the datamodels considerably.

A similar technique is proposed here for Grey Literature, to provide through a common catalog (based on a formal metadata datamodel), easy-to-use access to heterogeneous sources of grey literature.

CERIF

CERIF provides a comprehensive data model for R&D Information agreed by representatives of European countries (both European Union and associated States). The original CERIF1991 recommendation is being used as the basis for the ERGO Pilot Project (ERGO). The new CERIF data model has been produced within the CERIF Revision Working Group, dominantly by the author. The new CERIF datamodel is being implemented in UK as a RDBMS to ensure completeness, integrity and requirements satisfaction. CERIF assumes pre-existence of systems for patents, products and publications. However, CERIF has not (yet) tackled the problem of grey literature - hence the work by the author presented here. CERIF may be represented (Figure 1: CERIF High-Level Datamodel):

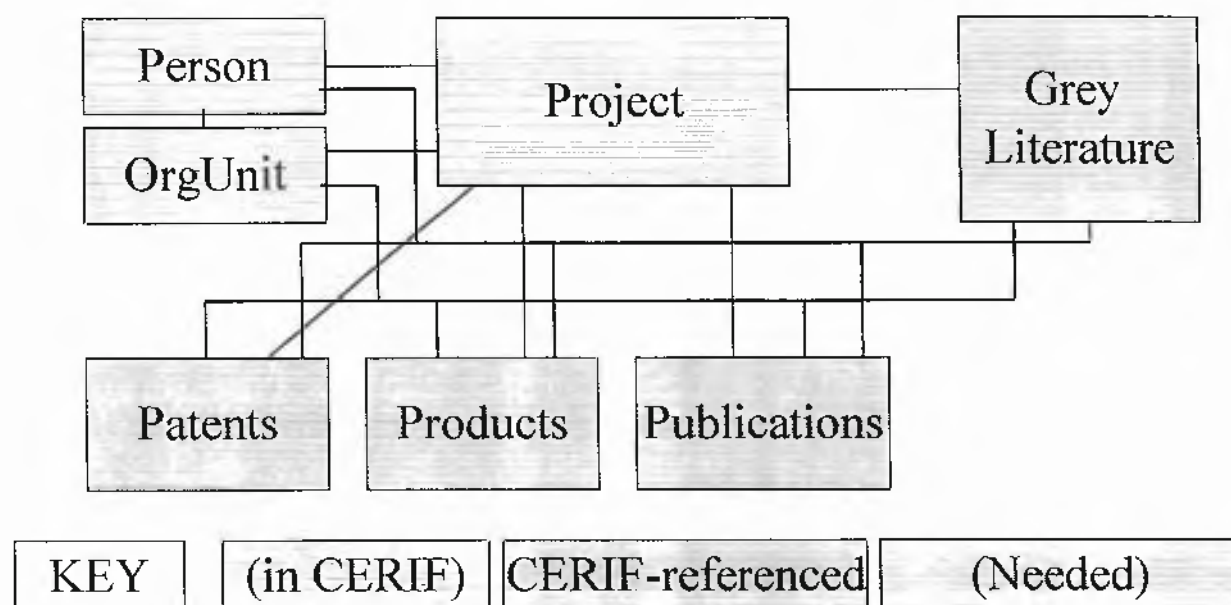


Figure 1: CERIF High-Level Datamodel

and from this 'ideal CRIS' (Current Research Information System) datamodel- which could be implemented in whole or part by anyone wishing to set up a R&D database system - a subset for exchange (respecting e.g. privacy legislation or commercial confidence) is derived and also a

datamodel for content metadata for use in a catalog. The full ERGO project architecture envisages end-user query to the catalog, from the 'hits' choice of relevant resources (objects) from the host databases and their provision via the exchange datamodel to the end-user as hypermedia documents.

GREY LITERATURE SYSTEM

Grey Literature System Requirements

The major requirements are as follows:

- (a) solicit / accept 'documents' produced in a variety of electronic forms (hyperlinked multimedia);
- (b) convert them to a canonical form for storage, including - where required - enforcing style (or store in original form with metadata to cause conversion);
- (c) register them and catalogue them in a suitable way;
- (d) provide suitable IPR, copyright and other legal protection;
- (e) make provision for charging for end-user services by end-user category and usage statistics collection;
- (f) provide a system for information retrieval over the documents (including multimedia);
- (g) make them available (subject to security / authorisation) to end-users in a variety of forms;
- (h) provide workflowed input / update

The 'documents' are:

- (a) as input: atomic documents / document fragments or hyperlinked document sets;
- (b) as output: atomic documents / document fragments or hyperlinked document sets

The key question then is what is the atomic unit? On input it is likely to be a complete document (author of complete opus) whereas on output it is likely to be the relevant fragment

of the document hyperlinked with other fragments of this and other documents to satisfy the user requirement.

Therefore it is essential to separate clearly:

- (a) document structure (hyperlinks between document atomic fragments);
- (b) document content (attributes or sections, language, character set, media representation(type));
- (c) document presentation / layout (for preservation of presentational features);
- (d) document metadata (e.g. cataloguing information, security / access, charging, provenance, certification..)

A Grey Literature Document is likely to be of text, graphics, and images and may include video and audio - but also with experimental results, virtual reality presentations - thus provision of 'drill-down' from the conventional view of a document to its source data will be essential.

GREY LITERATURE SYSTEM ARCHITECTURE

Simple Architecture

A simple architecture (where 'catalog' represents all required kinds of metadata) is:

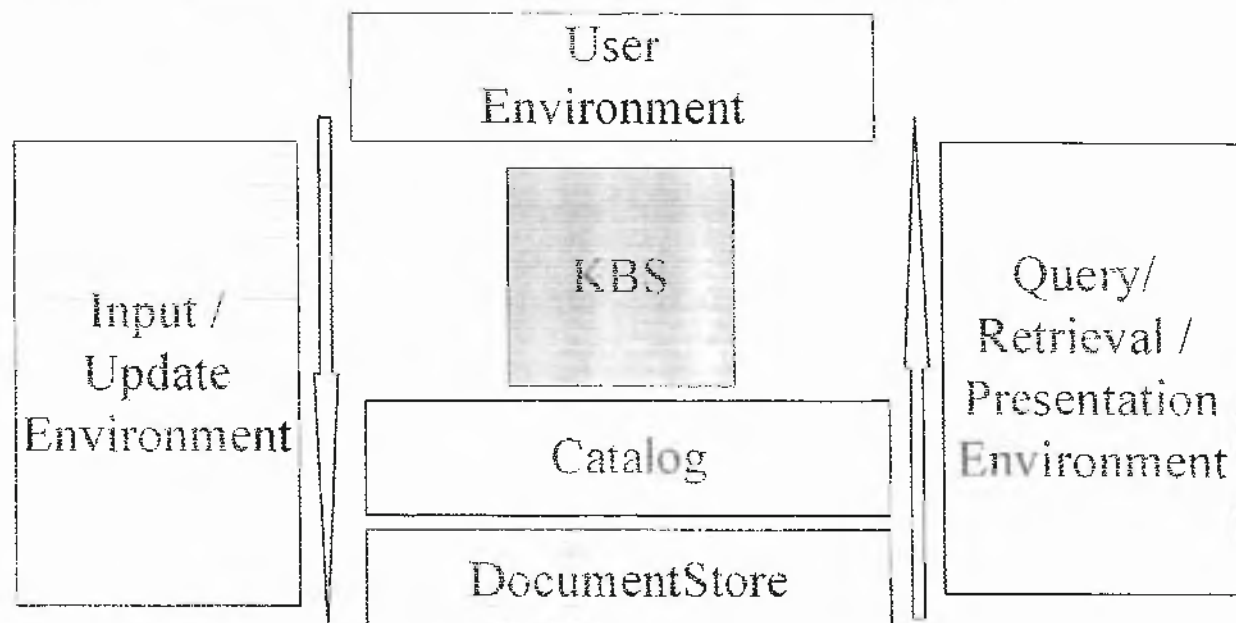


Figure 2: Simple Architecture

Major Architectural Ideas

The major architectural ideas are as follows:

- (a) accept and provide multiple forms - different formats (SGML, XML, HTML, Word...) and full multimedia (SMIL);
- (b) the use of rich metadata (of different kinds) not only as an index to the document content but also as a mechanism for security, pricing control, access information (royalty payments etc), provenance, certification - all linked to the Grey Literature processes implemented as workflow.
- (c) using KB assists and style controls to propose metadata catalogue entries having parsed the document or check input for quality,
- (d) the KB-assist knowing the metadata catalog and the user profile (implies domain ontology with the KBS) refines queries, improves integration of the answer - with appropriate explanation and controls presentation style for end-user preferences.

Architectural Features - Input / Update Forms and User Interface

The Input /Update Forms

The major idea is to use standard WWW HTML forms. However, using RDF-Workflow Technology (from the W3C-LA (World Wide Web Consortium - Leveraging Action) project (W3C-LA)) we can ensure that suitable authorisations occur for document placement in the library - e.g. author signature, supervisor signature, department signature and all with DSIG (Digital Signature) technology (for security) if required.

User Interface

The User Interface has to integrate at user client workstation access to all the required information sources, for both output and input as authorised, at the desk or in a mobile environment. It has to operate on various hardware / operating system platforms (Unix workstation, PC, mobile phone with display). There is only one solution for the architecture: use of a WWW Browser.

The browser should be lightweight and with a very fast core, and allowing add-ons and plug-ins configured dynamically as required. Experiments with such a browser have been performed at CLRC-RAL (for another project).

The browser should be supported by an easy-to use, tolerant, intelligently-assisting, multilingual, multimedia, multimodal user interface, characterised by:

- (a) easy to use : intuitive, minimal interaction;
- (b) tolerant : accepts incomplete and uncertain data (input and output) and proposes supplements to completion;
- (c) intelligently-assisting : query expansion / refinement - dialogue with user and answer expansion / refinement / explanation;
- (d) multi-lingual, multimedia, multimodal: all forms of human-computer interaction supported

Such an interface is to be implemented using Java Applets so that central control may be maintained, linked with evolving domain ontologies and multilingual thesauri, and compatibility with all reasonably modern browsers maintained.

Architectural Features - Metadata

There is relevant information on digital library metadata at the (UKOLN) site including pointers to the current standards and state of the art.

Content Metadata : The Catalog

This kind of metadata is equivalent to the conventional library card catalogue. However,

- (a) we have defined a content metadata database structure that is much more flexible than e.g. Dublin Core (DC) and is defined formally (Jeffery, Asserson, 1998a);
- (b) the design will use a structured DBMS (Database Management System) rather than an IR (Information Retrieval) system for fast retrieval, standardisation of query languages and optimal concurrency control, but especially to allow for flexible structural relationships both within content metadata and between content metadata and other kinds of metadata;

(c) the design provides for interoperability by providing interconversion from the catalog metadata to conventional Dublin Core stored in RDF (Resource Description Framework, the W3C metadata standard (RDF)); to multiple Dublin core extensions stored in RDF and to MARC (MARC) if required (as already demonstrated by the Nordic Metadata Project (NMP)) or to other metadata formats. This provides for maximum interoperability with other digital library projects yet does not constrain this design like the others.

Structure Metadata

The need to separate structure from content is well-known in the Hypermedia community by the use of a linkbase e.g. Hyperwave (HW), Microcosm (MI)). Unfortunately HTML was not designed with this in mind but mixes within the document hyperlinks (structure) and content.

However, there is a clear requirement to extend beyond the simple linkbase concept to accommodate links with semantics, ideally expressed as first order logic semantics with second order syntax (complex structures). This requirement and a solution has been demonstrated in interoperation of hyperlinked multimedia clinical patient records in the project Hypermedata (HMD). Such rich structure metadata implies deductive / inductive capability in the system - and thus the requirement for a KBS. For Grey Literature, as a first step, the use of extended RDBMS (Relational DBMS) technology is sufficient and provides an upgrade path to a richer system.

Presentation Metadata

This provides the recommended presentation of a complete document or document fragment for the appropriate stored form. Examples are : SGML - DSSL, XML - XSL. WORD - Stylesheet, full hypermedia - SMIL. Alternatively, documents can be stored in page-resolved form to preserve exactly the presentation required. Examples are PDF or Postscript.

Security /Access + Charging / Workflow Metadata

This kind of metadata provides appropriate information for processes associated with Security, privacy, authorisation, (copyright), with access records (charging / royalties), with document provenance and history - versions (link to workflow) and with document certification (e.g.

signed by author as original, signed by head of department as approved, signed by publication board as refereed (all linked to workflow).

Architecture - Detail

The above features lead to a more complex architecture (Figure 3: Detailed Architecture) while still preserving the essential features of the simple architecture (Figure 2: Simple Architecture):

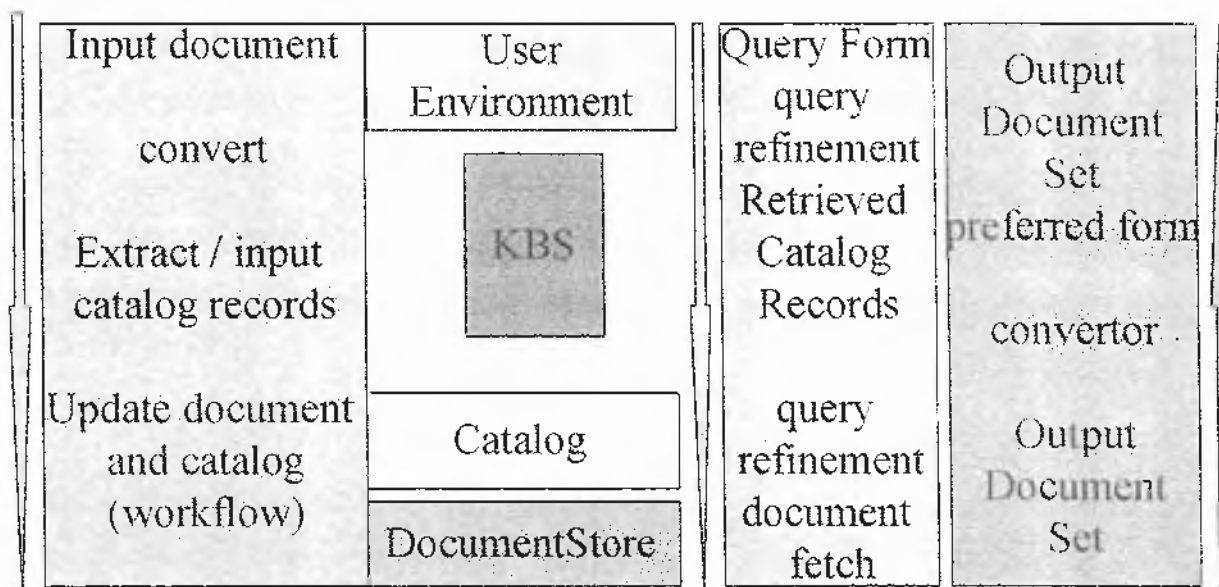


Figure 3: Detailed Architecture

Grey Literature System - Implementation

The architecture has been designed specifically so that it is possible to start simply and grow to the KBS assisted architecture starting with assists for the user interface and queries then to provide assistance in results and presentation. The implementation options have been left as flexible as possible: for example the KBS could be implemented as a central system (as in project MIPS (MIPS) or as cooperating agents e.g. as used at (Stanford).

The local process environment will dictate the need for workflow for authorisation, and for converters to handle different input and output forms in the environments for input / update (with workflow) and retrieval / presentation.

DISCUSSION

Implementing such an architecture raises several interesting implementation decisions which are discussed here.

Storage & Retrieval

- (a) centralised or distributed : multimedia assets are likely to be most accessed locally and are expensive to transmit so distributed is the best model. In practice logical central design and distributed physical design is ideal, but due to pre-existing systems unrealisable;
- (b) filestore or database store : the criteria are performance, security, and integrity. Database technology will be better for the latter two, performance will depend on several factors although there is some evidence that filestore systems perform slightly better than DBMS for Grey Literature Resources;
- (c) IR or Structured DB : the criteria are retrieval capability, update, and flexibility of structural linkage. A structured database - especially extended relational - will be better for almost all aspects, however, the one area where IR systems still excel is the use of inverted indexes for full-text searching. On balance, the advantages of Extended RDBMS outweigh those of IR systems.
- (d) same storage for metadata and data (objects, documents): there are advantages if both the metadata and data share a common environment for performance, security, integrity and a common DDL (data definition language - language to define schema) and common DML (data manipulation language eg SQL).

Presentation

The choice of presentation environments - and metadata structures to support them - is critically important not only for aesthetics but also for ensuring certain resources are preserved as presentationally intended for legal reasons.

The SGML Family: SGML: DTD, SGML, DSSSL, XML: DTD, XML, XSL and, of course SMIL provide matching content/structure and presentational capabilities and are clearly the best

choice with the richest (SGML) as canonical storage form; in practice it is likely that XML will be more widely used and this may well prove the best choice.

HTML: (DTD), HTML4.0, CSS2 provide some separation of content/structure from presentation but not separation of content and structure. Where it is not possible to solicit input in the SGML family or present using the SGML family then the HTML family is a widely available lower standard alternative.

However, there are many users of the Microsoft family of products: WORD: StyleSheet. However, the Stylesheet has rather limited presentational expression and so conversion from WORD to a canonical form is preferable.

Finally, there is the PageFamily: PDF, Postscript, proprietary DTP formats. Postscript should be used only as an interface between server and printer - there are too many variants for real compatibility. PDF provides a reasonably compact 'page image' form and is useful for preservation of intended presentation.

Input

There are advantages for cross-platform compatibility and maintenance (including downline loading of Java Applets) in using for input / update Web Forms. The optimal design decisions appear to be to handle the data so collected as follows: metadata --> XML; content--> XML or SGML or SMIL;

Alternatively, especially for large volumes of data it may be better to use a Context Sensitive Editor. In this case similar design decisions apply: metadata --> XML; content --> XML , SGML or SMIL

However, if a proprietary Word processor system is used e.g. Word, or a Text Processor e.g. LATEX or a proprietary DTP System then it will be necessary to apply converters - with success depending on the capabilities of the conversion software and the capabilities of the data generation systems to provide rich content and structure.

Metadata

One major architectural decision is that the key to success is metadata: it is used for

- (a) quality input : domain range checking, attribute value confirmation;
- (b) workflowed input process (including review, authorisation): ensuring integrity;
- (c) security, access control, charging, royalties: preserving the social and economic aspects of the system;
- (d) query assistance / optimisation : by use of the KBS interpreting the metadata of the domain ontology to improve the query formulation and to assist optimisation;
- (e) answer assistance / optimisation : by use of the KBS using metadata concerning presentation;
- (f) interoperability across heterogeneous sources : by use of the KBS and domain ontology providing schema reconciliation with metadata assistance in term clarification and domain comparison;
- (g) user interface adjustment / assistance : using metadata concerning the user profile and preferences;
- (h) help and explanation - education : using metadata to explain data content and structure, presentation and quality, and to explain process (such as workflowed) with associated rules and conditions.

An overview of Metadata for a R&D environment was presented at the CRIS98 Conference (Jeffery, 1998). W3C working groups are finalising RDF with XML implementation at the time of writing. Separately, the Dublin Core work appears to have stalled due to differences of opinion within the working group(s).

We have reviewed carefully content and structure metadata (that required for description of content and approximately equivalent to a library catalogue card) and have proposed a metadata datamodel which is a superset of all other content models and so they can be generated from it.

The datamodel is formal and yet flexible. Furthermore, it is extremely suitable for Grey Literature in a R&D environment because it integrates seamlessly with CERIF.

A current solution should be implemented on extended RDBMS technology since it provides adequate, if not extensive, intensional (schema) capability and sufficient formality. It also provides great flexibility for evolution towards more expressive environments (e.g deductive / inductive). The content metadata is detailed; other kinds of metadata are sketched and await further work in parallel with the work of W3C.

Our work leads to the following datamodel: (compared with the Dublin Core, deleted elements in *italics*, new elements in **bold**, all elements redefined)

content metadata:

UniqueId, Title, *Author or Creator*, **Person**, **OrgUnit**, *Subject and Keywords*, **Subject**, **Keywords**, Description, *Publisher*, *Other Contributor*, *Date*, Resource Type, *Format*, Resource Identifier, *Source*, *Language*, *Relation*, *Coverage*, **CoverageTemporal**, **CoverageSpatial**, *Rights Management*

note: each textual field has subelements language, representation, format; many elements have schemes; any element may be repeated. Thus this is normalised to:

<Unique^{1x}Id> <Resource Identifier> where the value of Resource Identifier is a pointer (typically a URL). /* links the URL or other pointer indicating the real resource (eg document) to a Unique Id in the metadata */

<Unique Id> <relation> <Person> where sub-elements of relation are type (meaning role), degree (cardinality) with mandatory / optional usage; associated relation constraints. Such relationships (roles) should have amongst their constraints temporal constraints including date or period. Sub-elements of person are a full attribute set from CERIF. Such relationships (roles) should have amongst their constraints temporal constraints including date or period.

<Unique Id> <relation> <OrgUnit> where sub-elements of relation are type (meaning role), degree (cardinality) with mandatory / optional usage; associated relation constraints. Such relationships (roles) should have amongst their constraints temporal constraints including date or period. Sub-elements of OrgUnit are a full attribute set from CERIF.

privacy metadata:

<Unique Id> <Privacy Scheme> <Privacy level constraints>

access metadata:

<Unique Id> <Access Rights Scheme> <Access level constraints>

charging metadata:

<Unique Id> <Charging Scheme> <Charge>

annotation metadata:

<Unique Id> <relation> <Annotation> <Person>

where <Annotation> has sub-elements language, representation, format. It is commonly necessary for comments upon a resource (annotations) to be recorded. The idea would be to define the purpose of the annotation by the type of relation (sub-elements of relation are type (meaning role), degree (cardinality) with mandatory / optional usage; associated relation constraints). Such relationships (roles) should have amongst their constraints temporal constraints including date or period) and for the annotator to be identified. In this case a DSig (digital signature) sub-element may be necessary under Person. This needs more discussion to determine the functional requirement.

These latter (non-content) kinds of metadata require further work (currently ongoing) to establish correctly the functional requirement, and then compare with emerging results of work elsewhere e.g. W3C. The approach taken here is much closer to the RDF (encoded as XML) approach of W3C - where essentially everything is reduced to binary relationships in a simple graph object model - than the existing Dublin Core work. A relational model of content metadata has been prepared and documented (Jeffery, Asserson, 1998a).

CONCLUSION

A joint project involving University of Bergen, University of Oslo (Norway) and CLRC-RAL (UK) has been initiated, following the Bergen Grey Literature Workshop, December 1998 (Be98) with the major primary aim of providing a Grey Literature System for research theses.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the work of colleagues in the CERIF Revision Working Group which provided part of the motivation for this work. The author also wishes to acknowledge the work of colleagues at the Bergen Grey Literature Workshop, December 1998, especially the organiser, Anne Asserson (who prompted this piece of work) and the team from University of Oslo led by Hana Konupek.

REFERENCES

(Be98) http://www.fou.uib.no/fou/grey_lit/

(BibSys) <http://www.bibsys.no/english.html>

(DELOS) <http://www.iei.pi.cnr.it/DELOS/>

(DC) http://www.oclc.org:5046/research/dublin_core/

(ERGO) <http://www.cordis.lu/ergo/>

(Farace,1990) SIGLE - System for information on grey literature in Europe Farace, D.J.. In: Login, vol. 14, no. 3, p. 6-9, 1990 - ISSN0920-0290

(FORSKDOK) <http://www.fou.uib.no/eng/>

(HMD) <http://www.itd.clrc.ac.uk/Activity/Hypermedata>

(HW) <http://www.hyperwave.de/>

(Jeffery,1992) Jeffery,K.G.'Database Integration' Proceedings SOFSEM92 Conference, Ed. M Bartosek, Masaryk University Brno, Czech Republic

(Jeffery.1998) <http://www.cordis.lu/cybercafe/src/jeffery.htm>

(Jeffery, Asserson,1998a) http://www.fou.uib.no/fou/grey_lit/literature.html (click on Data Model)

(Jeffery et al,1989) Jeffery,K.G.;Lay,J.O.;Miquel,J.F.;Zardan,S.;Naldi, F.; Vannini-Parenti,I.: 'IDEAS: A System for International Data Exchange and Access for Science'. Information Processing and Management 25(6) pp 703-711. 1989.

(MARC) <http://lcweb.loc.gov/marc/marc.html>

(MI) <http://www.mmrg.ecs.soton.ac.uk/projects/microcosm.html>

(MIPS) <http://www.itd.clrc.ac.uk/Activity/MIPS>

(NMP) <http://renki.lib.helsinki.fi/meta/>

(RDF) <http://www.w3.org/RDF>

(Stanford) <http://walrus.stanford.edu/diglib/>

(UKOLN) <http://ukoln.bath.ac.uk/metadata/>

(Virginia) <http://www.ndltd.org/>

(W3C) <http://www.w3.org>

(W3C-LA) <http://www.itd.clrc.ac.uk/Activity/W3C-LA>

Grey Literature in Energy: A Shifting Paradigm

by Deborah E. Cutler, U.S. DOE/OSTI

There were challenges to undertake and exciting new partnerships to forge as the exchange medium for energy information moved from paper and microfiche to electronic formats. As the paradigm for storing, retrieving, and disseminating energy-related grey literature has shifted and evolved, the Office of Scientific and Technical Information (OSTI), the Department of Energy's (DOE) information arm, has led the way. At this particular stage in the shift, OSTI has overcome a number of challenges and recognizes that there are still more to come. Dealing successfully with those challenges has come only as a result of a commitment to vision and cooperation/support from domestic and international partners. Of course the vehicles provided by the Internet, the World Wide Web, and related technologies were key, as well. OSTI has been able to offer the user community direct desktop access to a significant volume of energy-related grey literature. That access is FREE.

Producers and users of information have found themselves in a rapidly changing environment as the Internet and its support technologies have exploded. A companion event to that technology explosion has been the slashing of government budgets and dramatic downsizing of the government workforce. Government agencies can no longer afford to wait 10 years for proven solutions from private industry. There is an immediate need for answers to critical questions. How do you do more with less? What is the most efficient way to meet and exceed customer expectation? How can you keep commitments? How do you remove the "grey" from the traditional definition of grey literature? OSTI learned the answer to "how"—you take risks, you jump on the bandwagon, you just do it.

OSTI has been in the forefront and become a leader among US government agencies in offering grey literature via the Web. Evolving from the traditional world of paper and microfiche to the electronic world of PDFs and TIFFs has come quickly and not without growing pains. This paper documents OSTI's journey toward that new frontier and the challenges faced along the way.

THE OLD AND NEW FRONTIER.

Although the name and chain of command have shifted several times over the years, OSTI's core mission to collect, preserve, and disseminate scientific and technical information (STI) has remained the same. Over the last 52 years, OSTI's collection of STI has grown to over 1.5 million Energy R&D reports (traditionally viewed as grey literature) and approximately 5 million bibliographic citations representing worldwide energy R&D. The current database, today known as the *Energy Science and Technology Database* (EDB), has over 3.8 million citations. Over the past few years, OSTI has made significant strides into the information age: defining new electronic exchange formats; creating and linking to collections of digitized STI;

serving researchers directly; and developing a Web product called EnergyFiles: Virtual Library of Energy Science and Technology. As of September 1999, the DOE Information Bridge, a component of EnergyFiles, has over 56,000 full-text R&D grey literature reports (approximately 3.8 million pages), capturing a significant portion of DOE and global R&D output since 1996. The global information comes in through OSTI's partners in international agreements. To keep to the terms of these agreements, access to the full DOE Information Bridge (<https://apollo.osti.gov/dds>) is limited to the DOE and its contractor community. There is, however, public access to DOE's grey literature.

- ✓ The public version of the DOE Information Bridge (<http://www.doe.gov/bridge>) offered in collaboration with the U.S. Government Printing Office (GPO) Depository Library system, has over 43,000 DOE R&D grey literature reports (more than 2.8 million pages) that are electronically available and free to the public worldwide. OSTI is proud to have been among the first sites to make such a large volume of what used to be considered grey literature available to the Internet literate, winning several governmental awards for its efforts. Over the past 6 months, the two Information Bridge products combined averaged more than 2200 full text reports being downloaded weekly, far exceeding traditional distribution methods of the past. The grey literature is not simply available, but it is being USED.

The paradigm of grey literature being hard to get has indeed shifted for energy-related STI. So, how did we get there?

FIVE YEARS AGO

OSTI has always had responsibility for DOE's grey literature. For many years DOE and its contractors were required to provide paper copies of their reports to OSTI for public dissemination. Storage and dissemination at OSTI changed over time from paper, to microcards, and then on to microfiche. A bibliographic database citing the reports and other commercially published energy information was also created. Citations included abstracts and added subject indexing to facilitate retrieval of the information. Public dissemination of the DOE grey literature was accomplished through partnerships with the GPO Depository Library system and with the Department of Commerce's National Technical Information System (NTIS). Microfiche and paper continued to be the only ways to get this grey literature for many years. The constant challenge was to improve the timeliness of dissemination, because of the resource intensive nature of the storage and retrieval methods available to fill orders.

In addition to collecting and disseminating its own DOE grey literature, OSTI has a long history of international exchange of grey literature and related bibliographic information. OSTI's chief multilateral agreements are with the International Energy Agency Energy Technology Data Exchange (IEA/ETDE) and the International Atomic Energy Agency International Nuclear Information System (IAEA/INIS). ETDE, begun in 1987, currently has 18 countries exchanging the full scope of energy-related information (<http://www.etde.org>) while INIS, begun in 1970, has over 100 countries and 19 international organizations exchanging primarily nuclear-focused information (<http://www.iaea.or.at/inis/inis.htm>) From the inception of these agreements, OSTI has served as the US delegate and met US commitments as well as possible given recurring budget restraints. Since ETDE's inception, OSTI has managed and operated ETDE on behalf of its members as Operating Agent. As within DOE, five years ago the paradigm for grey literature exchange in these agreements was also paper and microfiche.

DRIVERS FOR CHANGE.

The 1990's of course brought better and faster computers, much wider use of the Internet, and a vastly more computer literate workforce. At the same time, reduced budgets and downsizing of government became the norm. OSTI in particular saw its budget decline by nearly half even though information and technology were continually recognized and touted as the new wave of the future. While new technologies are allowing for cost efficiencies, an up front investment is usually required. How to cope in this new environment with the new resource limitations became a formidable challenge. User expectations began to shift rapidly. Few would argue that the Internet considerably influenced the mind set of users who now want instant gratification. Information users today have little patience to wait weeks for a copy of a document; even waiting minutes for file downloads now seems like an eternity. The concept that because something is free it has less value has decidedly lost its influence. Many users today gravitate to what's easiest to get rather than the best or most scientifically correct. Bibliographic information alone is passe – users demand full text... Now!

To address resource issues as well as the new environment, OSTI knew it had to take advantage of new technologies and work with its partners wisely. It was no secret that more and more documents were being created in electronically shareable formats – as opposed to the sometimes cumbersome strict publishing formats of the past. Sites throughout the DOE community were evolving internally to new ways of producing their reports. Technology offered TIFFs, GIFs and PDFs as new options in the electronic viewing and printing of documents, without requiring users to have specific word processing software. Full text

searching was becoming a real possibility, bringing with it some schools of thought that bibliographic information would someday be obsolete. The vision, as OSTI saw it, was to give information to the user at the desktop, and this meant electronic full text.

In 1996, OSTI made a strategic decision to migrate to electronic full text and transition out of microfiche production. The challenge: what format(s) do you standardize on? Since DOE documents were continuing to come to OSTI in paper form, the format chosen for storage and manipulation was TIFF Group IV, and a scanning process was established. Microfiche production continued in parallel for some time, but was modified to use electronic TIFF images instead of the traditional paper and camera methods.

The next challenge was to choose a dissemination outlet. Thus the DOE Information Bridge concept was born. This Web concept had the bibliographic record retained as the base for searching but added a hyperlink to the full text. In addition, some full text searching capability was built into the product line. This was achieved by an automated OCR process of the scanned TIFF images to build a text index. Although no manual clean-up of the data was done, a good quality copy would produce quite good results, and served to enhance bibliographic searching.

align
p10. mfg
Price / 100.00
Another challenge faced during this time frame was whether access should be free. Should users see the information for free? Should they be able to download whole documents or only single pages? As a government entity, OSTI's work was already funded by taxpayer dollars. But this was added-value, and extra service. Then again billing meant accounting systems and additional resources. The end result was free access, and the capability for both page at a time viewing and full document downloads.

The next goal was to get partners involved. In collaboration with the DOE STI partners, regulations were revised to state that the preferred method of receipt of DOE reports would be electronic, and that soon dissemination too would be electronic. In early phases, only 5 formats were to be permitted on receipt: SGML, HTML, TIFF Group IV, PDF, and Postscript. The receipt format became known as the native format. At the time, SGML looked the most promising for scientific exchange of information, although experience has not borne that out. While OSTI does receive a few documents in SGML, the largest majority have been in PDF, followed by TIFF, then HTML, along with still quite a lot of paper. In recent times, OSTI has added acceptance of some word processing formats, but receipts to date have been few in number. TIFF images continue to be created from any paper received.

Regardless of the native format, OSTI initially chose TIFF Group IV as the standard for storage, dissemination and archival purposes. Over time, however, PDF has taken a dominant role, and PDF-wrapped TIFF images are also offered for dissemination for all documents residing on the Information Bridge. It is not clear yet whether OSTI will move to only storing PDF for archive, but it is a real possibility given the storage requirements to handle native, PDF and TIFF copies of each document. The challenge of bulk delivery of electronic full text was met by choosing the 8mm DAT as the output media for the TIFF images (in addition to their availability via the Information Bridge product line.) Short SGML files for each document accompany the TIFF images on DAT, to provide brief bibliographic information.

The next challenge was to involve international partners. OSTI's move to electronic full text certainly had impacts on these partnerships. OSTI worked closely with the INIS Secretariat in Vienna, Austria to share future directions and ideas, since US information made up a considerable portion of the INIS system. INIS also chose to migrate to electronic full text, with TIFF as the primary storage format following their own scanning procedures. However, INIS chose a CD-ROM companion product to the INIS bibliographic database as their dissemination media to members. With a number of countries in INIS still having limited Internet access, the CD-ROM option was chosen as the best choice to make full text available to all member countries. The main drawback of this system has been the number of CD-ROMs necessary to hold the data. INIS is currently looking at DVD options to condense the number required for a full collection, but much depends on the availability of the technology in those same countries where Internet is not a universal solution.

INIS has also chosen to transition out of microfiche dissemination, but decided to retain microfiche production for archival purposes, although it too is produced from TIFF images. In addition, OSTI receives an 8mm DAT of TIFF images from INIS to load as part of the limited access Information Bridge product, and provides INIS with the US full text information on DAT.

With OSTI as the ETDE Operating Agent, plans were shared within the ETDE community. ETDE member countries were encouraged to submit data electronically, and the Information Bridge and 8mm DAT dissemination were offered to ETDE input centers. Several countries receive the electronic full text to serve users in their country more directly. As Operating Agent, OSTI scans in grey literature in paper form received from ETDE members to create electronic full text. This is then added to the Information Bridge limited access product. While some documents are received at INIS and ETDE in electronic format, the large majority

continue to come in paper form. The last year has shown movement in some countries, however, to improve this situation.

Since 1996, many issues have arisen and have been met regarding electronic full text. Some that continue to be considered include which formats should be allowed, and which should be considered the archive. How can searching be improved? An issue related to ownership arose within the international community, that DOE was fortunate enough not to face. Some full text sources who were perfectly content with allowing microfiche copies of their documents to be distributed were not so eager to allow the proliferation of electronic versions. For some countries, this caused the unexpected side effect of a decline in the number of documents that could be made available by ETDE and INIS, making that literature even more grey in terms of ready access.

As electronic availability became more the norm within the DOE community, OSTI has offered DOE sites the option to send a URL to their documents, if the sites already have them on Internet. Within the Information Bridge structure, users can still hyperlink to the documents, but they will not physically reside at OSTI. Thus, storage of electronic full text is moving to a decentralized model even though retrieval continues to be facilitated by a centralized database. Archive issues remain a concern with this model, but sites are supposed to notify OSTI before eliminating their Internet access. Coming soon will be OSTI's implementation of persistent URLs or (PURLs) to allow more direct access to documents and to minimize the impact on users if documents are moved to a new location.

STILL TO COME?

For OSTI, new versions of the Information Bridge product line will soon be available, adding some enhancements for users. A popular request from users that will not be implemented right away due to cost is to make the full text searchable once downloaded. Neither TIFF nor the PDF-wrapped TIFF allow such functionality. While OSTI has looked into various options for doing so, all take time and resources.

EnergyFiles (<http://www.osti.gov/EnergyFiles>), OSTI's virtual library product that directs users to all types of energy information, continues to grow, with distributed searching options now being offered. In addition, a new product called PubSCIENCE (<http://www.osti.gov/pubsci>) debuts in October and will provide users with the capability to

search across a large compendium of peer reviewed Journal literature with a focus on Physical Sciences and Technology. Online access to the full text of many journal articles has been a much desired next step expressed by researchers. Although viewing of the full text from most publishers is not without cost, PubSCIENCE centralizes access to many journal articles without having to leave the office. In some ways, while journal literature has never been considered grey literature in the traditional definition, almost any paper-only product line could be considered grey literature in today's electronic environment.

Phy Sci
+
Tech

} change in
def!
of grey

In the international arena, ETDE has opted to have an Information Bridge-like product available to users in their countries beginning in late October. The product, to be called ETDEWEB or ETDE World Energy Base, will be offered from the ETDE Web Site at <http://www.etde.org>.

OSTI feels proud to have done it's part to meet the challenge of removing the "grey" from grey literature in energy by making it accessible. It has done so despite limited resources and with some difficult decision-making along the way. Without the commitment of a dedicated and capable staff and willing partners, grey would still be grey, and OSTI would be going down the path of the dinosaurs.

Contact information:

Debbie Cutler

International Programs Manager & IEA ETDE Operating Agent Representative

DOE/OSTI

175 Oak Ridge Turnpike

Oak Ridge, TN 37830

Ph: 423-576-1272

Fax: 423-576-2865

e-mail: debbie.cutler@ccmail.osti.gov

LINK MANAGERS FOR GREY LITERATURE

Elena Lodi, Dipartimento di Matematica di Siena

Martin Vesely and Jens Vigen, CERN

Abstract

In the self-service area of the library reading rooms it is necessary to organise the collections in the simplest way possible. This is an important feature for readers, making it possible to get direct access to the material without necessarily having to go via the library catalogue to retrieve the call number. A typical example will be that a collection is organised so that a reader can easily get directly from an article reference to the article itself in a library where the journal collection is simply organised alphabetically by title.

This requirement seems to have been forgotten by many of the most important actors in the digital library field, both by the commercial publishers and by many of the bodies producing grey literature. How are the users supposed to get directly to report xyz without having to navigate through n different web pages? This is happening as well in "The Library" which is intended to be close to 100% based on self service!

CERN Library has developed a mechanism called "Go Direct", to handle this problem. So far it only processes articles published in journals, but due to its success it will however be extended to include grey literature issued in series holding a report number.

The CERN Library catalogue uses "Go Direct" to automatically generate links from bibliographic information to the corresponding fulltext of a document.

The talk gives an overview of the present situation concerning link managers and URL architecture in the field of electronic publishing within high-energy physics.

Introduction

The CERN Scientific Information Service has for more than forty years been collecting and organising information of interest for the whole high energy-physics community. The aim is to have a collection as complete and up to date as possible. With the explosion of the Internet and the introduction of preprint servers the amount of information has multiplied several times over the last ten years, in parallel with the fact that user requirements are much higher today than ever before. Our Service is now processing around thirty thousand new grey research documents per year in addition to "whitewashing" thousands of other documents collected in earlier years.

With the introduction of electronic preprints (Dallman et al., 1992) followed by electronic journals (Chaney et al., 1999a), completely new possibilities for weaving the information together have arisen and so have new problems. In the pre-electronic era it was possible to get directly from a reference to the corresponding document in the library, often without using a catalogue, while now the same operation would require a significant amount of navigation through a number of web pages. The CERN

Scientific Information Service's main aim in organising the information has been to avoid unnecessary navigation and to introduce new identification systems. Related documents are linked together and the system permits the users to get to a given document simply based on the bibliographic reference without having to go via hierarchical trees or any other "clicking exercises".

The very first solution to this problem was to have a simple form with a script behind it which, based on the submitted information, could compute the URL of the required document. The idea was to provide a facilitating tool for the scientists so that it should be easy to retrieve the articles available online. As the form, now called "Go Direct", became successful, it also became clear that the script could instead take its arguments directly from the library catalogue so that links could be created "on the fly".

Collection and manipulation of information

Most of the acquired documents in the CERN Library which are considered as grey literature originate from the Los Alamos Eprint Archive, but there is also a substantial amount of documents originating from inhouse CERN research and direct submission from other institutes. The collected information, the metadata, is unified and presented within WebLib, the web interface to the library catalogue and various other CERN databases. This interface enables easy access to the metadata and other features such as ordering documents, loan renewals and access to the fulltext for a significant amount of the documents.

As soon as a document is introduced into the system, it is automatically analysed. The references are extracted from the file and stored separately in order to allow later citation searches. The extracted results are then processed through a "normalisation filter" in order to ensure that the linking program will recognise the references. This means that correctly analysed references, referring to online available documents, will actually point the users directly to the corresponding fulltext.

Most scientists prefer to read the published version of an article when it is available, instead of the preprint. Consequently an important role of a library handling masses of grey literature is to connect such documents together as soon as the published versions appear. This procedure in the CERN Library is handled fairly automatically for documents published in the journal literature by a program that queries WebLib with the intention of matching and updating the preprint records for the documents which were published.

For each article, depending on the source of information, different metadata will be available. Normally at least the three first authors will be supplied, the year of publication and the published title in addition to the publication reference itself. The database index is such that the answer to a query is faster if the search is done on the basis of consecutive words. For this reason the program selects the longest phrase of consecutive words from the title. A word must contain only letters. Moreover some words, like the logical operators *and*, *or* and *not*, are stop words. A phrase is a consecutive sequence of words separated by spaces. For example in the title *Hadronic ZZ, W⁺W⁻, and production with QCD corrections and leptonic decays*, there are 3

phrases: namely *Hadronic ZZ, production with QCD corrections* and *leptonic decays* of length 2, 4 and 2 words respectively. The program will then choose the phrase *production with QCD corrections* as an argument to be supplied to the matching algorithm.

A query submitted to WebLib cannot be longer than 74 characters and cannot contain more than 7 logical operators. The program makes the query using all the available authors, the longest phrase of words in the title and the year 199?, which covers all the years in the range [1990,1999]. If the answer to the query returns one or more records, the titles are examined by a pattern matching algorithm. The problem of pattern matching is the following: given two strings called a text T and a pattern P , if P occurs in T , P is a substring of T . Let T equal to $\langle X_1, X_2, \dots, X_m \rangle$ and P equal to $\langle Y_1, Y_2, \dots, Y_n \rangle$, where $m \geq n$. P will then be a substring of T if there exists a value i , $1 \leq i \leq m-n+1$. Consequently the string $\langle X_i, X_{i+1}, \dots, X_{i+n-1} \rangle$ is equal to P when $X_i=Y_1$, $X_{i+1}=Y_2$, ..., $X_{i+n-1}=Y_n$. For $T=\text{Parallel updating cellular automaton models of FK systems}$ and $P=\text{cellular automaton models}$, P is a substring of T starting from position $i=19$. If P is equal to T the problem is called perfect matching (Crochemore and Rytter, 1994).

Approximated pattern matching (Bertossi et al., 1990) is an extension of the pattern matching problem which allows for the possibility of having errors in the matching. A posteriori the most common errors of typing are the following:

- a mismatch, in the sense that the two corresponding characters in P and T are different
- a missed character in T
- a missed character in P , that is a character more in T ,
- inversion of two consecutive characters.

For example let $T = \text{Non-universal Quasi-Long Range Order}$ and $P = \text{Nonuniversal, Quasi-Long-Range Ordre}$, P is recognized as a matching of T with 4 errors: a missed character in P ($X_4 = "-"$), a missed character in T ($Y_{13} = ", "$) a mismatch ($X_{25} = " " \text{ different from } Y_{25} = "-"$) and an inversion ($X_{35} = Y_{36} = e$ and $X_{36} = Y_{35} = r$).

In the program there is also a treatment of formulae. In WebLib formulae are given following the syntax of LaTeX which means that they will be delimited by \$ signs. Superscripts and subscripts are indicated by ^ and _ respectively and their values are within parentheses, usually { } Many of the information sources used for the updating eliminate all these special symbols from the titles, consequently the matching program is also using the "linear" version of the title too. For example the title = *Search for η_c' and h_c (1P_1) states in the $e+e-$ annihilations*, will appear in Latex as *Search for $\eta_{c'}'$ and h_{c1} (1P_1) states in the $e+e-$ annihilations*, becomes typically *Search for etac' and hc (1P1) states in the e+e- annihilations*.

When a preprint is accepted for publication it will normally be revised when going through the referee process and the title itself might even change. The most frequent changes in the title are added words. The program finds matches on the basis of single words, takes as pattern the shortest title in words and allows missing words. For example *From the Hubbard to the SO(5) ladder: a numerical study* is recognized as a substring of *Evolution of the low-energy excitation spectrum from the pure Hubbard ladder to the SO(5) ladder: a numerical study*.

Finding matches on single words makes it possible to find abbreviations as the following example shows: *Six-body LFTD and wave functions for the massive Schwinger model* is recognized in the title *Six-body Light-Front Tamm-Dancoff approximation and wave functions for the massive Schwinger model*. The program counts errors on the basis of missing words. For example, if the title furnished by the information source is *Chiral odd structure functions from a chiral soliton* we can find the following three titles in WebLib:

- Nucleon structure functions from a chiral soliton
- Structure functions from chiral soliton models
- Chiral odd structure functions within a chiral soliton

All these three titles are at distance 1 from the given text. In fact when the program analyses the first title, it finds all the words in the pattern except the first one, *Nucleon*. In the second case the missing word is *models* and in the third *within*. A

human being can decide quickly that the third title is the one to choose, but the algorithm can not. For this reason the database is automatically updated with the references to the journal, volume, issue and page only when so called perfect matches are found. In the other cases the approximate matches are written into a file that will be further analysed by a librarian. The work of the librarian is facilitated by an additional program that displays the bibliographic data of the article on the right side of the screen, and on the left side the data of the matching candidates. The program also provides links to the fulltext so that the librarian, if necessary, can compare the actual texts.

The matching program is written in Perl and has more than 4000 lines of code. It takes only one parameter, a list of published papers or simply a table of contents, (let's call it INDATA) and produces 4 files called `verify.indata`, `noconfidence.indata`, `correct.indata`, `append.indata`. The first and second files are created in the same directory where the program is located. The file `verify.inadata` is a file of verification to check if errors occur during the processing of the data. The file `noconfidence.indata` is the one used for the treatment of approximate pattern matching. The files `append.indata` and `correct.indata` are later run in batch mode to update the catalogue.

Automatically linking to the fulltext

Traditionally a preprint was simply retrieved in the library collection by its report number in the same way that an article was retrieved by its triplet journal title, volume and page. Moving into the electronic age this was paradoxically not as simple

60 Direct *

any more. In most cases the users were obliged to point their browsers to different preprint archives or publishers sites and start navigating from there, even if the reference of the document they were looking for was already known. CERN scientists wanted, knowing the reference, to get directly to the fulltext of the documents. The whole project started in 1997, not a long time ago but still in the early days of electronic journals in the CERN Library (Chaney et al., 1999b). Looking at electronic journals, at that time only Physical Review D, published by the American Physical Society (APS), had chosen to publish using an URL architecture making it possible to associate a bibliographic record with a fulltext article without having to store the URL in the database (Doyle, 1999). The term "link manager" hardly did not exist. Since then a long period of lobbying other publishers to follow the example of APS has followed. Surprisingly enough this has not been straightforward, the approach has apparently been too pedestrian for many of the important players in the field. Today it is however adopted by the Institute of Physics (IOP), EDP Sciences (EDSP) and Elsevier has implemented it for a few important titles in the field of high-energy physics. In addition there are publishers which can be handled by maintaining a set of look-up tables at the CERN servers.

Having thousands of publication references in the library database it became clearly interesting to automatically locate the online published articles using Internet technologies, especially as articles were also made available retrospectively, predating the electronic preprint archives.

The electronic journals are organized in remote databases within the publishers information systems and are accessible via the Internet restricted to the subscribers. This implies the need to introduce some Internet-based mechanism that would provide the connection between the bibliographic information available from the library catalogue and fulltext of the corresponding document.

Such a tool has been developed, aiming to serve the WebLib which can be accessed and browsed from distant terminals and search a variety of databases independently of their location. This combination became a very comfortable tool for document research. However, since no identification standard for electronically issued documents has been established, the online fulltext version of published articles could not be located in a trivial way.

Recently the topics of identification of electronic documents have been discussed widely. The need for a unique identifier lead to several de facto standards for identification, i.e.: Digital Object Identifier (DOI), Publisher Item Identifier (PII) or Serial Item and Contribution Identifier (SICI) (Doyle, 1999). For the location of a document on the Internet the given reference needs to be transformed into a URL through which the document is accessible. The URL is in fact not a permanent identifier and some standardized way of referring to online documents should be established. However, such a standard has not yet been accepted by a majority of the community. In addition to that, the well established system for referring to printed material, using the triplet journal title, volume and page, is still around and will

continue to be for many years. It appears therefore natural for the time being, to build an identifier based on the traditional standard reference.

Before the link manager was implemented, the most obvious solution was to store the particular link in the library catalogue and whenever the full text was required, the URL was available directly from the catalogue. This way the catalogue records contained both the standard reference and the URL itself. By introducing another field in the bibliographic record, this solution was easy and straightforward. On the other hand, the amount of articles that had to be handled required too much manual maintenance. Additionally URLs are not stable and can be freely modified at any time. The update of the modified URLs would be even more demanding, even if logically possible, because it would require periodical checks throughout the entire database. The acquisition of metadata could have been automated by building an engine that supplies the bibliographic records to the catalogue with a unique URL for each article. Eventually, the development of this engine would be much more powerful embracing a different approach to the solution.

The idea is to not store the URLs in the catalogue at all, based on the fact that the URLs are not considered as identifiers and therefore they might be variable. The valid URLs are then composed at the very moment the requests arise. In order to be able to perform this composition the system includes another independent module: the link manager. The link manager is a service based on web technology that provides applications with valid links to specified documents online. The idea is to keep the

links up to date by moving the responsibility for their validity from static storage in the catalogue database to a resolution mechanism. Unlike the database, this resolution mechanism does not store any links, but creates them dynamically according to the current configuration. In case the final URL changes during its lifetime, the link manager has only to be reconfigured. The link that appears in the client application is then replaced with the address of the link manager that composes the final URL from the given reference.

The mechanism consists of two parts that cooperate as necessary (link builder/indexing robot), where one takes care of solving the requests and the other one maintains the metadata repository. The metadata repository contains firstly the configuration that defines how the URL will be designated in particular cases – for certain documents or sources. This configuration is the fixed and required input for the link manager to work. Secondly, in case of handling non-standard references, the system needs to index some additional information, which can be gathered using the indexing robot. This indexing has to be performed periodically as a batch process, usually triggered by date. The link builder then has three inputs: the reference asked by the client, the configuration for the particular source and the indexed information.

The link manager module can be used by any application that has to deal with linking to fulltext documents. As mentioned before, besides WebLib the link manager is used for the form "Go Direct" that allows the user to request articles without going via the catalogue.

In the WebLib solution (fig. 1) the link manager provides the connection between the bibliographic information available from the library catalogue (local database) and the fulltext information provided by the publishers over the Internet (external database). The range of Internet searched locations is significantly restricted. The robot goes only through those locations that are defined in the metadata repository as bibliographic information resources - in this case the subscribed electronic journals.

The link manager resolves the standard reference passed through WebLib and returns the valid URL. No URLs are stored in the library catalogue or in the metadata repository, but are created "on the fly". When invoked from the web application the link manager automatically switches to this determined location so the document can be retrieved.

The link manager has to be flexible enough to be able to link to many different types of locations. The system retrieves data from different publishers and all of them support different strategies for data storage and therefore different URL structures. The download of metadata and its indexing has only to be performed when the actual URL contains information that cannot be derived from the standard reference. In many cases this can be the issue number within a certain volume, the internal document number assigned by the publishing house, your client identification or any other information that for some reason or another can be used as an identifier.

A few publishers have their own link manager service, which simplifies the URL designation on the client side, because the system only has to deal with the standard supplied by the publisher. In this way is it possible to build some generic algorithms dealing with the particular standards. In all other non-standard approaches special algorithms have to be defined.

The URL then contains the location of the particular link manager supplied by the publisher and the standard reference. Looking at an example, the APS link manager (<http://publish.aps.org/linkfaq.html>) locates the article in the following way:

[http://publish.aps.org/abstract/\\$journal/v\\$volume/p\\$Id](http://publish.aps.org/abstract/$journal/v$volume/p$Id)

where the following variables must be supplied:

\$journal abbreviated journal title

\$volume volume

\$Id starting page

The URL for any APS article, generally presented, will then look like:

<http://<link manager address>/<formatted standard reference>>

which then passes the request to the external resolution machine.

Similar for the non-standard references:

`http://<server address>/<directory location>/<fulltext file>`

which links directly to the article.

In the final state the system provides the connection using two resolution machines, one on the client side and one on the publishers side (fig. 2). The local link manager allows the unification of the records stored in the library catalogue, enabling access to many different sources. It is also desirable to choose the identifier independently of the external environment. In this way is it possible to use the standard reference (the triplet) also as an identifier for online versions of the electronic documents, without introducing other fields in the catalogue. The local link manager then transforms the defined identifier to the standard that so far depends on the agreement between the two communicating sides. The remote link manager will eventually resolve the final location of a document that is being retrieved. The "two link manager solution" simplifies the process of linking to the fulltext of any article by defining the interface on the application layer.

Systems with a link manager only on one side can also be implemented. The local system will then have to take over all the work and it therefore becomes significantly more complex. Such a solution requires building up a quite large metadata repository

in order to also be able to handle the non-standard references which of course is more difficult to maintain. If the link manager is always available remotely, the local link manager is in fact outsourced and to switch between different sources, a simple CGI script can be then used instead.

At CERN some documents are provided directly by the local information system. Therefore the system in fact enables quite flexible two-way exchange of scientific information by both linking to external resources and allowing other institutes to link directly to local fulltext documents.

Future challenges

The developments which were meant to make the world simpler for the scientists made it in some senses actually more complex. The documents are in principle easier to get access to, but there are more references to a single document and its published counterpart than there were ever before: author/title, report no., preprint archive no., URL, DOI etc., etc. and finally the traditional bibliographic citation - or at least it is still there for the time being. The goal must be to reduce these different references to a minimum and establish a standard which is as intuitive as possible. Having made the major publishers aware of the power of the link managers and later getting several of them onboard, the situation seems to be turning in the right direction. The work of lobbying the publishers, also the ones issuing grey literature, is however not yet at an end. To augment the pressure the campaign should also be directed to other database

providers. The more people applying this simple system, the more likely it will be that it will become a de facto standard. The major challenge for the future will be however to find a better solution for how to associate grey literature with its published counterparts. The DOI Foundation, a foundation more and more publishers are joining, is just aiming for this. The DOI is meant to be a unique identifier of any piece of intellectual content, together with a system for using that identifier to locate digital services on the Internet associated with that content (Paskin, 1999). In the field of high-energy physics there is already a veritable preprint culture which is accordingly well organised. More or less all these preprints have report numbers and it is not difficult to imagine that they accordingly could be assigned DOIs. Exploiting the power of the DOI system, one could then reach a situation where by pointing to the DOI of a preprint, the researcher would automatically be offered the option to follow a link to the published counterpart as soon as this would be available.

References:

- Bertossi, A. A., Lodi, E., Luccio, F., Pagli, L. (1990) "Two string matching problem with weighted errors", *Theoretical Computer Science*, Vol 73, pp. 319-328
- Chaney, E., Bulliard, C., Christiansen, C., Cressent, J. P. (1999a) "Une bibliothèque de recherche face à l'édition électronique: l'exemple du CERN", *Bulletin des bibliothèques de France*, Vol 44 No 2, pp. 27-32

Chaney, E., Bulliard, C., Christiansen, C. (1999b) "The electronic journal service at CERN, a first evaluation: user access interfaces and user awareness", *Vine*, No. 110, pp. 23-29

Crochemore, M., Rytter, W (1994) *Text algorithms*, Oxford University Press, New York

Dallman, D., Draper, M., Schwarz, S (1994) "Electronic pre-publishing for worldwide access: the case of high energy physics", *Interlending & Document Supply*, Vol 22 No 2, pp. 3-7

Doyle, M (1999) "Citing and linking in electronic scholarly publishing: a pragmatic approach", edited by Smith, J. W. T., Ardö, A., Linde, P., *Electronic Publishing '99. Redefining the information chain – new ways and voices: proceedings of an ICCC/IFIP conference, Karlskrona/Ronneby, Sweden, 10-12 May 1999*, ICCC Press, Washington D.C., pp. 51-59

Paskin, N. (1999) "The digital object identifier system: digital technology meets content management", *Interlending & Document Supply*, Vol 27 No 1, pp. 13-16

Figure 1

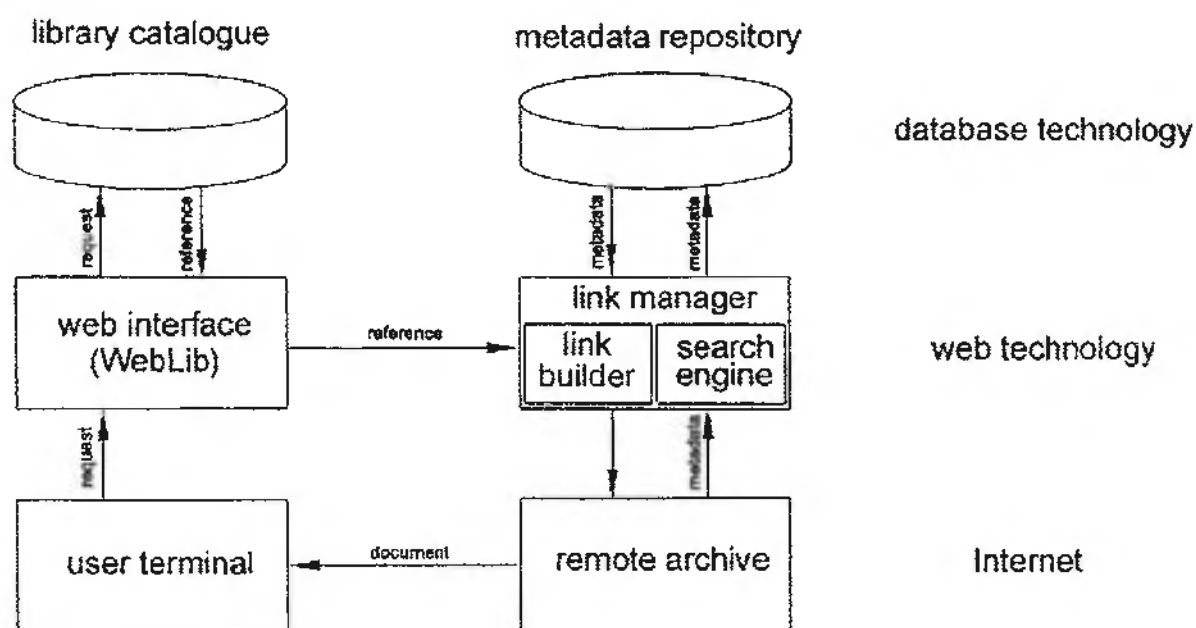
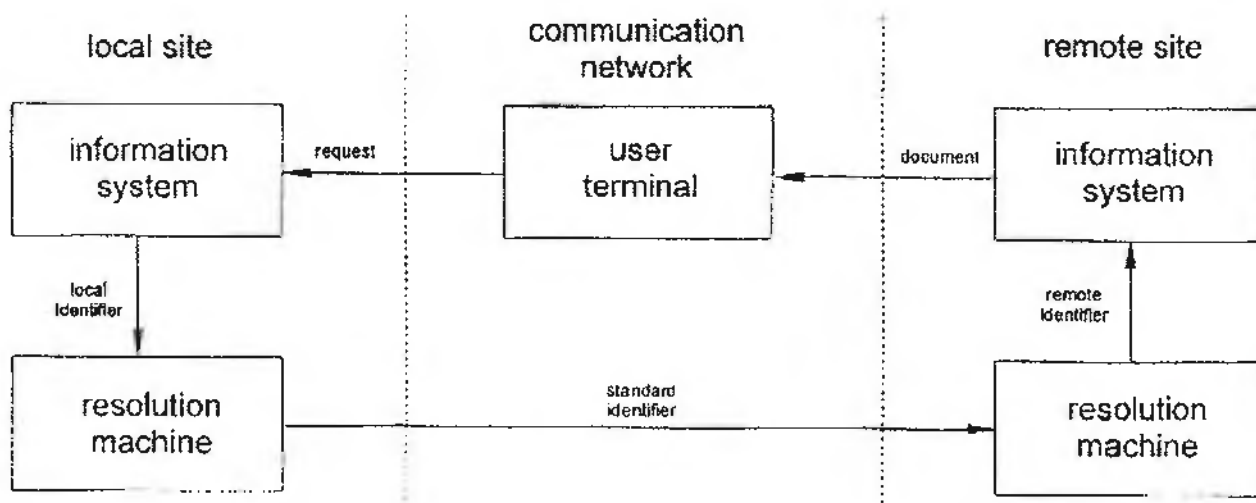


Figure 2



Grey Copyrights for Grey Literature: National Assumptions, International Rights

AUTHOR

Michael Seadle
Digital Services and Copyright Librarian
Michigan State University
100 Library
East Lansing, MI USA 48823-3533
Editor, *Library Hi Tech*

ABSTRACT

Copyright enforcement is a mixture of law and culturally-influenced practice. Although the statute law appears in black and white, the traditions and expectations of those who create and use gray literature are definitely grayscale. And the latter are a better predictor of both infringement and litigation than the letter of the law itself. Casual US assumptions about using gray literature stem in part from technicalities in the pre-1978 copyright law, which often threw gray literature immediately into the public domain. These assumptions stem also partly from the low economic valuation of gray literature, and the key role economic value plays in US copyright litigation. Since moral rights plays almost no role in US copyright law, common European features such as the right of withdrawal have no place in local expectations. In recent years digital publishing and international access have exposed these assumptions to a context where the separation of moral and economic rights is taken for granted. Legal rights, both within the US and Europe, have also come in conflict with the urgent desire of both scholars and librarians to make gray documents more accessible by republishing them on the Web.

[ARTICLE BEGINS]

INTRODUCTION

"This, no doubt, is why natural history, in the Classical period, cannot be established as biology. Up to the end of the eighteenth century, in fact, life does not exist: only living beings." -- Foucault, 1973, p. 160

It would be easy to paraphrase Foucault and say: until the end of the 1960s, grey literature did not exist, only uncontrolled documents. Grey literature as a concept is relatively new. Everyone realized that some documents did not get into standard bibliographic tools, did not appear in library catalogs, did not, in fact, have the status of "real" publications from legitimate commercial presses. An archivist might squirrel them away in Hollinger boxes with a cryptic label and a note on a typed finding-aid that probably listed them under the name of the person or organization who was peculiar enough to collect them. They were the untouchables of the library world, a permanent low caste because they served purposes tainted by self-promotion, private

interest or practical application. They had, in the most literal sense, little or no value. They were not even bookstore outcasts, since they never once reached the shelves.

Copyright issues for grey literature need to be understood in this human and historical context. The law itself is no different for a mimeographed pamphlet run off in blue ink by a volunteer organization than for a hardcover book with crisp, stylish printing, and a professionally designed dust-jacket. But the attitudes toward these different works, both for creators and consumers, diverge because of the relative value assigned to each. When the physical object appears to be worth more, the intellectual property tends to be treated accordingly.

Property rights depend in part on the willingness of people to defend them. The owner of a piece of land who allows public passage across it may in time lose the right to sue for trespass, perhaps even to deny access. The same is true for intellectual property rights, especially in the United States, where, until 1989, notice and registration requirements were an integral and important part of the law. The author of a work of grey literature had to care enough about it to comply with the notice requirements to have any hope for protection. Many did not. In fact, many intentionally did not because they saw copyright only as a restriction on the dissemination of information that was their real object.

The tensions between law and practice are illustrated below in three examples. In the first, the creators of the grey literature deliberately ignored copyright requirements in order to improve access to their message. In the second, the rights holders claim has come under attack because of initial irregularities, and their rights have been hard to enforce. In the third, copyright protection was unambiguous and was actively claimed, but open, blatant infringement still occurred. To understand these examples, a brief overview of US copyright law may be helpful.

AN OVERVIEW OF US COPYRIGHT LAW

This overview covers five issues relevant to grey literature: 1) notice and registration, 2) duration of protection, 3) published versus unpublished works, 4) moral rights, and 5) enforcement.

No one should read this overview as legal advice, or even as an accurate portrayal of the full range of complications of copyright law in the US. Those who wish to examine the statutes themselves can find them on the Web courtesy of Cornell University Law School. [1] Those who wish to learn some of the basic case law may wish to consult a compilation like West Publishing's *Cases and Materials on Copyright*.... (Nimmer, 1991)

Notice and Registration

The notice requirement went through three historical phases. The first lasted until the major 1978 revision of the copyright law. In it, notice in the form of the word "copyright," the abbreviation "copyr," or the symbol "©" was a prerequisite for protection. Omit it, however accidentally or unintentionally, and the work, if publicly distributed, fell immediately and irretrievably into the public domain. The second phase lasted from 1978 until 1989. The law still required a notice for published works, but allowance was made for minor accidental omissions, which could be rectified without permanently losing protection. The final phase runs from 1989 to the present. Under it, the notice requirement has been abolished in order to bring US law into accord with the Berne Convention. Of course notice applied only to text. The status of spoken words, broadcasts, and performances, though generally protected, was more ambiguous until 1978.

The registration requirement, like the notice requirement, became less absolute after 1978, but remains necessary for some types of damages. Post-infringement registration is also possible. There is a \$30 registration fee – low enough not to be impossible for the poor, though perhaps still high enough to daunt the frugal.

Duration of Protection

Before 1978, the US law protected works for a maximum of two 28 year terms. The second term was not automatic until 1964, which meant that many works received only the first 28 years of protection. In any case the absolute maximum was 56 years. After 1978, the term of protection followed the Berne Convention standard of the life of the author plus 50 years, or 75 years from publication for anonymous works. This meant that accurate knowledge of the death date of the author became more important in determining the term of protection than the date of publication, which could generally be found even in a grey literature work. In 1998 the period of protection increased to life plus 70, or 95 years for corporate or anonymous works.

Published vs Unpublished Works

Before 1978, unpublished works did not come under the federal copyright statute at all. They were governed instead either by state or by common law. The latter gave unpublished works perpetual protection, with no need for a copyright notice, no need for registration, no need in fact for anything except that the works be fixed in a relatively permanent form. This meant that letters, personal papers, and any other works which never met the definition of publication had enduring protection. Works which had a limited distribution among a private audience, such as the draft of an article or book, could well (and generally did) fall into this category.

After the new copyright law came into effect in 1978, unpublished as well as published works received protection from Federal law. A grace period prevented unpublished works from falling into the public domain until the end of 2002 at the earliest (if created before 1978) or 2047 (if created after). Except for these limits, the usual rule of life-of-the-author plus 70 years now applies to all published and unpublished works (95 years for corporate and anonymous authors).

Moral Rights

Moral rights for text or phonorecords have no place in US law. The US Congress did enact limited moral rights protection for the creators of works of visual art as part of the bargain for joining the Berne Convention. But none of the extensive set of European rights that allow an author to prevent defacement or to withdraw a written work exist in any form in the US statutes. Like the Anglo-American tradition generally, the US copyright law focuses almost exclusively on economic rights.

In Germany and France, moral rights are inalienable. The creator cannot sell them or give them away, though they can be inherited. They are also personal rights, and a corporation cannot own them. No such restriction exists in US law. In fact, the trend toward corporate ownership of copyrights seems to be growing.

Enforcement

Rights owners must choose whether to enforce their privileges or not. It is neither cheap nor easy to enforce the copyright law against a determined infringer. A private letter to an infringer may work, but most rights owners start with a letter from a lawyer. If the infringement continues, pursuing the issue means going to court. It also means having to register the work, pay the fee, deposit at least one copy of the work, and fill out a two-page form. The court process

is slow, and a large corporation can easily drag out the proceedings while legal costs build up for an individual.

Winning a court case does not necessarily result in a windfall. The statutory damages are limited to \$100,000 for willful infringement (17 USC 504c). This could be bankrupting for an individual, but trivial to a corporation. Recovery for actual damages may be higher, but requires some proof of what those losses were, which can be particularly difficult in grey literature case, because of market value uncertainties. (17 USC 504b) The copyright owner can also recover any profits made by the infringer, but the court can take into account expenses and elements of profit not attributable to the infringement, which easily becomes an accounting nightmare.

EXAMPLE ONE: IGNORING COPYRIGHT

In the late 1960s a political party known as the Black Panthers invaded campuses and television screens with their message of black power and independence. They wrote a wide variety of pamphlets to help spread their message. One of these was Eldridge Cleaver's "Credo for Rioters and Looters." (Cleaver, 1969). Another was Huey Newton's "Essays from the Minister of Defense." (Newton, 1967) These were small works, just a couple of pages long, typed, duplicated, definitely not printed on any sort of professional press. Party faithful handed them out on street corners, no charge. They represented a classic type of grey literature: self-published, free, intended to get across a message not make a profit. No one called them grey literature, though. Propaganda was the preferred term.

Copies of both of those pamphlets found their way to the Special Collections department of Michigan State University Library. This was somewhat unusual. They were nothing like Darwin first editions or the papers of a prize-winning author. It happened, however, that MSU Libraries houses the Russel B. Nye Popular Culture Collection, which includes American Radicalism. The pamphlets found their way into a vertical file, were cataloged, and became popular with students doing research for the freshman writing class called American Thought and Language. Most of the rest of the copies of those pamphlets went into waste baskets shortly after distribution. They had value to almost no one except the party faithful and scholars building a collection for the future.

In 1968, the operative copyright law in the US was still the old 1909 law. Copyright protection was not automatic, and even registration of the work meant nothing if all the copies did not contain a copyright notice of copyright. Leave off that notice, and any work which met the definition of being published fell immediately into the public domain, from which no rescue was possible.

No evidence exists for how the Black Panther Party, or Huey Newton, or Eldridge Cleaver actually analyzed the situation. But three suppositions seem reasonable. First, that the amount of trouble to register the pamphlets in accordance with the law seemed out of all proportion to the benefit. Second, that they actively wanted them copied, stolen, distributed as widely as possible, which copyright protection would only hinder. And third, they did not believe in the US government anyway. The latter was an outgrowth of their political views and not typical of grey literature publishers generally. The other reasons may have been more broadly shared: the Popular Culture Collection certainly has a large number of other pamphlets from the same and earlier eras without the required copyright notice -- all from organizations that explicitly just wanted their message to get out.

In a sense, the decision to eschew copyright did, in the end, benefit these two Black Panther pamphlets. Michigan State has been able to scan them and mount them on the Web in

digital form for students to use in research, and for anyone in the world with an internet connection to see. They have outlived the Black Panther Party, and their message is more widely available than ever.

If the Black Panther Party had published the same two pamphlets today, the legal situation would be very different. The lack of a notice and registration for the works would not matter. They would have automatic copyright protection. In fact there is no way that they could lose their protected status, whether they wanted to or not. Protection is both automatic and inescapable. If they wished people to have the freedom to reproduce and redistribute the pamphlets, they would have to take the proactive step of including an explicit permission in the work. In other words, it would require some minimal knowledge of US copyright law to achieve the same open access.

Although the Black Panther Party is gone, plenty of other groups are willing to pass out handbills, pamphlets, and other forms of grey literature similar in quality and quantity. This is particularly true around election time, and anyone who has seen the phenomenon has probably also seen the residue in nearby litter baskets. For this type of grey literature, the current copyright law has done nothing to increase its likelihood of retention. The message would go to the landfill, or to a paper-recycling plant, with perhaps one or two copies that a foresightful librarian files away in an acid-free folder in a vertical file waiting for the 95 year term for corporate, or the life-plus-70 term for authored works, to expire before access can legally be increased.

At present under US law, neither Newton nor Cleaver or their estates would have the moral right to withdraw the pamphlets because of changed views. Their choice to publish their views became immediately irretrievable.

EXAMPLE TWO: BELATED COPYRIGHT

On August 28, 1963 in Detroit, Michigan, Dr. Martin Luther King gave a speech which is now called the "I have a dream" speech. Not everyone classes speeches with grey literature, but they share many grey literature characteristics, especially when a text of the speech is distributed. King's speech is better known than the thousands of others that took place during those active years of the US civil rights movement, but it was not sold and not published by a standard commercial or university press. According to OCLC's WorldCat database, only about 15 libraries own the newsletter in which the text was distributed. Considering that this speech quickly became one of the most famous, and perhaps the most influential, in post-World War Two America, the scant listing is astonishing.

The speech and the newsletter copy ran into copyright problems almost immediately. King wanted a large audience for his speech. He encouraged the press to cover the event and to broadcast it nationally. Under the 1909 Copyright Act, King's speech counted as performance, not publication. King also had a history of registering his speeches for copyright protection, and in fact sent a copy of the speech to the Copyright Office for deposit on September 30, 1963. But the organizing committee had asked him to provide an advanced text or summary for the press. The copies distributed to the press had no copyright notice. Because of its impact, the speech was widely repeated and reproduced, including by a company called Mister Maestro, Inc. In a 1963 decision of the Southern District of New York, the court disagreed with the argument that the press copies had thrown the speech into public domain. It held that only a limited distribution of copies took place, not one which would constitute publication. The court recognized King's copyright, and assessed damages. (Nimmer, 1991, p. 159-167)

Nonetheless copies of the speech continued to be made and distributed, even though the King estate was relatively diligent about trying to enforce its rights. Few other owners of grey literature copyright have similar moral and monetary resources to fight infringements, though few grey literature examples are equally coveted. CBS was among those who included part (62%) of the speech in a recent documentary videotape, which they sold. Although the estate took them to court, a district judge reversed the prior ruling in 1998 and decided in favor of CBS, because he felt the circumstances surrounding the speech did in fact meet the definition of general publication. The ruling has been controversial and is currently under appeal. (Ringel, 1999)

However the appeal turns out, this example illustrates the kind of legal complications that can arise from the ambiguous publication status of a work of grey literature. Today the same set of problems could not be repeated under US law, because of the abolition of the notice requirement. But other similar ambiguities, such as unclear work-for-hire situations, exist. Ambiguities in general are more likely to involve grey literature because the urge to get information out, as with the King speech, often overrides any desire to make sure economic rights are protected.

In any case even the 1963 decision upholding King's copyright did not stop willful copying. People continued to act as if the speech were in the public domain.

EXAMPLE THREE: EXPLICIT COPYRIGHT

Although desktop publishing software and laser printers made grey literature vastly easier to create in the 1980s, the growth of the Web in the 1990s turned almost every college student at a major university into a potential grey literature publisher, since most students have access to enough network-accessible storage to mount HTML files, and perhaps even a few scanned images. Much of this Web-based grey literature is the most ephemeral sort. It appears and disappears at the whim of its creator. Some of it may be for a particular class and vanish at the end of the semester. In any case universities tend to close the accounts within a year or so of graduation. Some students migrate their writings to a new home page on a commercial Internet Service Provider. But even when the contents remain unaltered, the address has changed.

One attractive example of this form of grey literature is Erica Olsen's home page "Es Chica Chica." [2] She set it up as an "ezine," and after several attempts, persuaded Yahoo to register it under "Entertainment > Humor > By Topic > Job Humor > Librarians" as "Why you Should Fall to your Knees and Worship a Librarian." Does registration with Yahoo make a site less grey? Perhaps, but the other problems of transience remain.

Unlike many Web publishers, Olsen has been careful to put copyright notices on her pages. She works in a research library, and has been accepted at library school, so she has a better than average understanding of copyright law. She knew that the law no longer required her to put the copyright notice on her Web page to give it protection. She added it only to warn visitors that she cared about protecting her material. (Olsen, 1999)

Her site has been popular, especially the three paragraph long "Why you Should Fall to your Knees and Worship a Librarian" section, and she maintains a guestbook for comments. In July, 1999, a visitor wrote:

Your website ROCKS!! We must have made 20 copies of "WHY YOU SHOULD...WORSHIP A LIBRARIAN" and passed them around to the 40+ librarians in our system. Gal, you're a welcome addition to our profession - come work in beautiful Long Beach, CA! -- McMillan, 1999

The visitor not only admits a violation, but seems to think that the infringement will please the copyright holder. Since the visitor appears to be a librarian herself, it may be reasonable to assume that she has a greater than average awareness of copyright law. She may well have behaved differently if Olsen's three paragraphs had appeared in a traditional print journal or in a regularly published book. But her attitude toward at least this form of grey literature seems to be that the copyright owner wants as much exposure for her message as possible. Her reasoning could easily be: access to the home page is free, so why should Olsen care about copies?

Olsen has told friends and colleagues about the infringement, but did not contact McMillan, although she had her email address. And she has not considered any legal action. This is typical of many grey literature infringements. The infringer probably really meant no harm, complaining would not undo the (largely moral) damage, and a lawsuit seems out of all proportion to her means or to the infringement itself. Instead, Olsen plans to be careful about putting any of her writings on the Web that she might later sell. (Olsen, 1999)

In practice, copyright offered no protection, even when handled with foresight and care.

COPYRIGHT, GREY LITERATURE, AND THE NEXT MILLENNIUM

Change in an ecology is systematic. When one element is changed, effects can be felt throughout the whole system. Local changes can disappear without a trace if they are incompatible with the rest of the system. -- Nardi, 1999, p. 51

Will the casual attitudes toward grey literature, and infringing practices that accompany them, disappear in the new millennium? Two relatively recent events have changed the grey literature's ecological niche in potentially significant ways. One was the long-awaited US membership in the Berne Convention at the end of the 1980s. The other was the commercialization of the Internet in the mid-1990s. The former, in effect, extended protection to all grey literature produced in the US, and the latter created tens of thousands of virtual publishers who could market access to information without traditional publishers, bookstores, or any of the conventional bibliographic apparatus. At the least, Web publication has blurred the dividing line between grey and non-grey literature. Access to grey literature on the Web is currently not much different than, and often easier than, access to for-fee publications.

Web-based publishing has brought with it greater international access. This has exposed US grey literature to people in societies where the existence of both moral and economic rights is taken for granted. This exposure may ultimately increase respect for grey literature copyrights and reduce the readiness to infringe, especially if the philosophical basis for US copyright protection becomes less exclusively associated with economic value. But the process seems likely to be slow, in part because it conflicts with the urgent desire of scholars, librarians, and the creators of grey literature themselves to maximize accessibility. Unfringement probably will not decrease until the creators decide to defend their rights. Until then, at least in the US, copyright for grey literature is almost meaningless.]

NOTES

[1] <http://www4.law.cornell.edu/uscode/17/>

[2] <http://www.msu.edu/user/olseneri/>

REFERENCES

- Cleaver, E. (1969), *Credo for Rioters and Looters*, Black Panther Party. Available: <http://www.lib.msu.edu/spc/digital/radicalism/e185.61.c551969.htm>
- Foucault, M. (1973), *The Order of Things: an Archaeology of the Human Sciences*. NY: Vintage.
- McMillan, S. (31 July 1999), Message 226. Available: <http://www.guestbook.de/yasg.cgi?X=142268&P=2>
- Newton, H. (1968), *Essays from the Minister of Defense*, Black Panther Party. Available: <http://www.lib.msu.edu/spc/digital/radicalism/e185.61.5n481968.htm>
- Nardi, Bonnie A., and Vicki L. O'Day, *Information Ecologies: Using Technology with Heart*. MIT Press, Cambridge, MA, 1999.
- Nimmer, M., P. Marcus, D. Myers, D. Nimmer (1991), *cases and Materials on Copyright; and Other Aspects of Entertainment Litigation, including Unfair Competition, Defamation, Privacy*, 4th edition, West Publishing, St. Paul, MN.
- Olsen, E. (8 September 1999), conversation with Erica Olsen, East Lansing, MI.
- Ringel, J. (17 May 1999), "King vs CBS," *Connecticut Law Tribune*. Available on Lexis-Nexis.

Intellectual property on the move

Some observations on authors' rights, grey literature, publishers and last but not least access to information in a cyberian age

drs. Cees A.T. de Blaaij*

Introduction

The purpose of this article is to shed some light on the legal developments regarding intellectual property and grey literature in the digital environment.

First of all I want to discuss the current legal changes that are taking place in the copyright field and the electronic environment we call cyberspace, the electronic superhighway or any other overworked term which may refer to the conception of a computer networked world. Secondly I want to address the issue of authors' rights and grey literature in a networked information space.

I shall try to answer the question in what ways the current developments in copyright law will affect the electronic publishing and the access to grey literature. Also I would like to go into the role of commercial publishers and libraries in this matter.

Pandora's Digital Box

This age we call the information society is the outcome of a transitional process from an industrial economy evolving into an information based economy. To put it more adequately: in this process of transition we make in general assumptions about the relevance of information often in the terms of the market and individual ownership. Copyright challenges not only our notions of property but our basic assumptions about the circulation of information in terms of production, distribution, accessibility and exploitation.

The introduction of digital technology and the coming of computernetworks linked to each other throughout the world created the fertile environment for fast and easy reproduction and almost unlimited distribution of an electronic work. These works include all kinds of copyright protected writings, images and sound recordings; all of them having one thing in common: the electronic format, patterns of ones and zeros.

This electronic format has the fluency properties of water and like a river heading to the sea electronic information flows from the creator to the enduser. However for some creators – e.g. those working in the music recording industry – the digital environment nowadays has become the modern version of Pandora's box because of its facilitation of unlimited reproduction and distribution of copyright protected digitized material all over the world.¹

In this way the Internet has become a celestial jukebox for downloading music, software programs etc. for free without any consideration or - perhaps - knowledge about intellectual property rights. For copyright holders – like the entertainment industry, software producers etc. - the continuous infringement of their copyrights was a signal for starting a lobby campaign to protect their interests in the digital world. With some success. There was a lot of

pressure on legislative bodies for introducing and implementing new copyright laws tailored to the needs of copyright holders active in cyberspace.

Berne Convention and the TRIPs Agreement

The Berne Convention is the foundation of contemporary international copyright law dating from 1886 and since then there have been many revisions². Most countries did sign this convention. Some rather late like the US who signed the Convention in 1988. Most national copyright laws are based on the Berne Convention.

In 1994 the World Trade Organization (WTO) Agreement was signed together with an Annex to this called TRIPs – the Trade related aspects of Intellectual Property rights³.

Things gained momentum with the acceptance of the TRIPs Agreement which pushed the economic value of intellectual property by offering developed nations a more effective regime of protection in return for greater access to their markets on the part of commodity producers. But at the same time these producers also happened to be the consumers of products which were subject of intellectual property protection. The TRIPs Agreement strengthened in general the protection of intellectual property by supporting restrictions on freedom of access to knowledge and information.

Of course it is a strange combination when on the one hand the freedom of trade is stimulated by breaking down trade barriers and on the other hand restrictions on the access to knowledge and information are enforced.

The WIPO Copyright Treaty

With the World Intellectual Property Organization Conference in December 1996 there were a number of far reaching proposals on the extension and expansion of copyright which were finally rejected in the adopted version of The WIPO Copyright Treaty⁴.

The WIPO Copyright Treaty is a special agreement under Article 20 of the Berne Convention⁵ and grants to authors more extensive rights than the Convention. The rejection of the original draft was of great importance because the draft was heavily weighted against public and user interests.

The proposals regarding to the right of reproduction and the right of communication in Articles 7⁶ and 10⁷ of the original draft WIPO Copyright Treaty were moved forward just a few months before the Conference started. This narrow time schedule reduced the opportunity for national countries to give comment and consult with user groups for reaching a kind of consensus.

Right of Reproduction

If article 7 had been accepted then this would mean that all electronic copies no matter how temporary were reproductions. In practice this would make copies made in a computer memory or shown by means of a browser on a screen an infringement unless permitted by law or by the copyright holder.

Instead of article 7 an agreed statement was accepted which stated that the reproduction right as formulated in Article 7 did fully apply in the digital environment. But at the WIPO conference no binding decision was made regarding the reproduction right which covers temporary copies.

Storage on a hard disk or in a computer memory is a particular mode of temporary or transient fixation. But the definition also depends on the type of technology used. And as we all know technology is dynamic by nature. For example ways of storing information by means of caching may become outdated because there will be other and more sophisticated solutions for storing electronic information. This means also that original article 7 in this case would become obsolete.

This shows even more that we don't know enough from the technological problems we may encounter in the future. Therefore, from a technological point of view it would be premature to implement the original article 7.

The adopted WIPO Copyright Treaty grants to authors more extensive rights. An extension of the period for protection of works was adopted to life plus 70 years. In my opinion no real arguments can be found for justifying this adoption. It is really no incentive for authors to produce works that otherwise will not be produced. The extension benefitted the publishers because the extension resurrects copyright in some works after the old period of protection has expired.

The adoption of the treaty was heavily influenced by an intensive lobbying campaign from the side of large commercial publishers and information providers. They pushed for a copyright reform centered on the extension of copyright protection and the expansion of the concept of copyright. This position is being dictated by the fear of what unforeseen consequences may follow in a world of fast changing technology.

This is also visible in the U.S. Report on the National Information Infrastructure⁸ and the E.C. Green Paper on Copyright and Related Rights in the Information Society⁹. Also the European Commission made a recent proposal for a Directive¹⁰ which results in the strengthening of the position of copyright holders interests.

Libraries and researchers oppose that transient copying is infringement because they want to apply the same rights of users which are valid in the analog world. Publishers on the other hand fear that approving temporarily stored electronic information will open Pandora's digital box. From their point of view it will be better to protect any type of temporarily stored information to prevent any erosion at all. This is what I call the Fort Knox¹¹ approach. But the question is still open and has to be resolved within the context of Article 9¹² of the Berne Convention if temporary copying constitutes an infringement.

The Digital Millenium Copyright Act

The implementation of the WIPO Copyright Treaty in the US has resulted into the introduction in October 1998 of The Digital Millenium Copyright Act¹³ (DMCA) to meet the conditions which were formulated in the WIPO Copyright Treaty.

This new addition to American copyright law is very complex by nature. The DMCA gives the greatest protection to copyright owners' right to control access to information since it make it a violation both for users to circumvent access controls, and for others to manufacture, distribute or offer devices that circumvent technological measures as means of protection of copyrights.

The DMCA leaves some uncertainty for the public interest because Congress delayed implementation of the prohibition on circumvention of access controls for a two-year period¹⁴ and further instructed the Librarian of Congress in consultation with the Register of Copyrights, to identify particular classes of works whose users who would be "adversely

affected by the prohibition...in their ability to make noninfringing uses under this title of a particular class of copyrighted works¹⁵. It is not clear if these classes of works include scientific journal articles and other materials like grey literature which are relevant to non-profit purposes.

In general you get the impression that the DMCA in its final appearance is systematically incoherent. All those involved in the negotiations got a piece of the copyright pie but in a way that resulted in securing certain extremely specific exceptions, rather than mentioning general limiting principles. They did not look or rather failed to look at the broad picture. As a result it is uncertain how the balance between owners and users will evolve in the future. Taking the developments in European Union into account the overall picture doesn't seem very bright for users and libraries .

The Nature of the Copyright Balance

Another important question is raised in what ways constitutional rights are harmed with these new forms of digital copyright protection. There may be a point that it may be necessary to bring copyright in balance with the U.S. First Amendment and Article 10 of the European Convention on Human Rights.

I refer here to the U.S. Constitution. In Article 1, § 8, clause 8 confers upon Congress the power to "promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." The copyright protection is derived from this clause. The author is granted exclusive rights by means of the Copyright Act to promote useful arts and sciences by securing for the life-time of the author plus 50 years of several exclusive rights with regard to their writings.

So the exclusive rights only exist to the extent they are necessary as an economic incentive for creation. There is no legal or moral basis in the American Constitution for authors to have complete control over the works they create¹⁶.

If we look at copyright history then we see that this thought already originated within The Statute of Anne, passed in England in 1710 and a landmark in the history of copyright law. It recognized that authors should be the primary beneficiaries of copyright law; it also established the idea that such laws should have only limited duration, after which works could pass into public domain. The designated period came to be set at 28 years.

So what it is stated within the US Constitution is not new. Congress made it ninety years already clear that the interest in protecting the rights of authors is secondary to the interest in maintaining public access to creative works¹⁷. Also article 10 from the European Convention on Human rights guarantees this type of freedom of communication and information.

The exclusive rights from the copyright holder exists as a tool and are necessary as an economic incentive for creation. So the conclusion is that there must be rights sufficient to provide an incentive to create new works but not to an extent that hinders access to works. Both in the European Union and in the US there is a constitutional basis for this conclusion.

Keeping the Balance

Without paying enough attention to the rights of public access and use and at the same time not reaching a general consensus in society, the new developed copyright legislation therefore serves the already established few who can get them passed and does not benefit society as a

whole.

In the preamble of the WIPO Copyright Treaty there was a Recital which recognized the need to keep the balance between the rights of authors and 'the *larger* public interest', particularly education, research and access to information. In the original draft Treaty there was no such Recital. So it is important that the Recital exists and confirms the primacy of the public interest as a goal to the dissemination of knowledge and free access to information and ideas. It is necessary that a better tuned legal framework has to be developed that can be used to balance the proprietary and commercial interests with all parties involved.

As already stated before, the larger public interest is constitutionally protected in international jurisdictions by a guarantee of freedom of communication including the right to disseminate ideas and information and have access to them.

Intellectual property law contains numerous balances between the right of creators against the rights of the public at large in a rich public domain but overprotection of copyright holders – the Fort Knox approach – won't benefit either user nor creator in the digital world. In the words of judge Kozinski:

"Overprotecting intellectual property is as harmful as underprotecting it. Creativity is impossible without a rich public domain. Nothing today, like nothing we since we tamed fire, is genuinely new: culture, like science and technology, grows by accretion, each new creator building on the works of those who came before. Overprotection stifles the very creative forces it supposed to nurture"¹⁸

The right of access to information in the electronic environment is just as fundamental as protecting the intellectual property rights of copyright holders.

A new direction ?

The digital environment makes it necessary to adapt existing copyright law to new technological circumstances but one can ask if traditional intellectual property ideas and values from the offline, physical world will also work in the legal field online world.

According to John Perry Barlow¹⁹ - one of the founders of the Electronic Frontier Foundation²⁰ - we have to do with a profoundly new kind of challenge.

In his view the copyright law from the physical world cannot be applied to a digital environment because in his words "the accumulated canon of copyright and patent law, was developed to convey forms and methods of expression entirely different from the vaporous cargo it is now asked to carry"²¹.

Should we as Barlow observes it 'dance on the grave of copyright'²². This will solve very little according to Barlow 'especially when so few are willing to admit that the occupant of this grave is even deceased and are trying up by force what no longer can be upheld by popular consent'²³.

In my opinion Barlow's prophecy is a too premature in announcing the death of intellectual property. History proves that old media are never fully swallowed up by newer media so intellectual property will be there to stay, even in the digital world. Notwithstanding Barlow's judgment on the future of copyright his arguments show us that we need a different approach how to deal with Pandora's digital box.

Compared to the off-line world the networked environment creates totally new possibilities for communication and information distribution all over the world. Information has become a

commodity in an economic world where breaking down trade barriers and the promotion of free world trade is the underlying driving force.

As a kind of case study for new models of exploitation rights I will examine the situation of grey literature in a networked space and intellectual property issues related to this.

The case of scientific e-publishing

There are two types of grey literature which have an important share in the area of grey publishing in general. On the first place scholarly information and secondly government information²⁴ in different kinds of format: working papers, policy statements, reports etc. A fairly new development is the substantial amount of grey publishing currently carried out on the Internet by a number of scholars and researchers without the involvement of commercial publishers.²⁵ The way Internet offers new possibilities made many people realize that the commercial model is not very comfortable to the need to have free circulation of ideas among scientists. And that is what is essentially about: scientific communication is a matter of sharing ideas and have free access to a common pool resource.

Typically enough this is the second time this happens in history because the birth of the scientific journal itself in the XVIIth century was also an answer to the needs of scientists to have a communication platform to stay in touch with each other.

Mutatis mutandis the electronic journal and e-print archives also pick up the basic idea of sharing information and scientific knowledge. And this comes very close to the objective of copyright as formulated in US Constitution, the European Convention of Human Rights and more specifically in the Recital of the WIPO Copyright Treaty pointing at the digital environment.

Commercial scientific publishing

Commercial publishers entered the field of the scientific journal from the 19th century onwards. After the Second World War, commercial publishing became the dominant mode of distribution. Nowadays only a small number of international publishers like Reed-Elsevier, Wolters-Kluwer, Springer and several others have established themselves as the most important players in the business of distributing scientific journals.

Profits and prices

Commercial publishing has become big business. The mergers and acquisitions which have taken place and continue to occur with great frequency are but a reflection of what's happening in the business and industry. The large publishers go for market share, industry penetration and profits. It is no secret that Reed-Elsevier enjoyed high returns on its stock. Reed-Elsevier reported profits of \$ 378 million on sales of \$ 938 million in its scientific activities alone. It is also no secret that Reed-Elsevier sold its consumer magazines publishing division so it could focus on scientific/technical journals with a higher profit-margin. Reed-Elsevier also wants to be the dominant publisher in scientific/technical journals market. Some fifteen hundred titles are listed in the current Elsevier Science: 1999 Subscription Price List. Prices range from a modest \$ 6500 for an annual subscription to *Analytica Chimica* to a nice price of *Excerpta Medica* for \$ 49000. Subscriptions to all journals would cost between 1.6 million and three million a year and it still goes up.

Commercial publishing competitive and value-adding ?

Another question is if the commercial publishing industry is competitive enough as a mode to produce at the least costly mode. If we look at the amount of profits of some commercial publishers then this remains to be seen. Many feel that publishers of high-price, small circulation journals are making excess profits. Can somebody justify the doubling of the cost of Elsevier's Brain Research between 1992 and 1996 to \$ 15,000 annually ?

What can fairly be expected ? On basis of earlier arguments I would welcome an investigation to anti-competitive practices in the scientific publishing sector from the US Government or in the European Union. The reason for this is on the first place to get a balanced view on the whole matter of scientific publishing and secondly to correct the threatening monopolization of scientific information by stimulating competitive practices in scientific publishing. So in the end we can pay a reasonable price for information in an area where information needs to be shared between users according to the formulated constitutional principles of scientific information.

Do high subscription rates also justify the quality of commercially published scientific journals ?

The production process of a journal consists of authoring, editing, presentation, duplication, distribution, promotion, integrity assurance and branding. Each of these activities may contribute to a higher value-added product and are essential features of paper publications. Are they worth the cost ? The cost consist of the amount of money involved in producing the journal and the delay in publication creating a loss of currency. So the only way to judge the value would be to enhance the competition between modes of publishing to reach an affordable price level.

Serial crises

Why do they charge so much for what is essentially a printing and distribution service ? The answer is simple. They can do it because the market allows it. Prices rise to the maximum level the customer can bear with catastrophic results for academic libraries in general. The US Association of Research Libraries calculated that 114 member libraries spent 142 per cent more on journals in 1997 than ten years before, but ordered 6 per cent fewer titles. This is what is called the 'serial crises'.

Role of libraries

Many academic libraries, at least in my country and many other countries in Europe, could not cope with the phenomenon of steep rising prices and canceled subscriptions to a lot of scientific journals. This way of economic exploitation of scientific information has lead in some cases toward a situation of information poverty, in which rights of accessibility for students and teachers are violated in the perception of scientific information as a public good. In this way a fundamental relation exists between the freedom of information and the exploitation of intellectual property rights.

Publishers like the traditional journals because take care for a steady income flow. For libraries it is a totally different world. They are left to the mercy of those who eat their budgets, so it is quite natural from the perspective of survival that libraries cancel subscriptions to particular journals, while on the other hand making sure that some other library in the same area continues to subscribe. In this way sharing arrangements are made. The cancellation of subscriptions reduces the income for the publishers and 'forces' them – they say – to increase the subscription prize. This produces the countereffect: libraries will

continue to cancel subscriptions. So is there any hope left for libraries ?

The library 'empire' strikes back

In my opinion the end is approaching in what libraries can endure. Libraries in the US and Europe are going to strike back, the ultimate goal being to return control of scholarly publishing to the non-profit societies and what they consider to be responsible publishers. Then again this supports the idea of a more free circulation of scientific information.

Can the circle be broken ?

Feasibility scientific publishing: the pioneering role of grey literature

Is scientific publishing on the Internet without the involvement of commercial publishers feasible?

As stated before more competition between modes of scientific publishing is welcome as a matter of lowering costs and at same time reach an acceptable level of quality.

The way how grey literature is being produced and published on the Internet could serve as a prototype how things might work successfully for electronic scientific publishing in general. The publishing of grey literature in the digital environment sets an example for other categories of electronic publications.

In my opinion even in the paper society grey literature was more successful in communicating novel ideas mainly through direct communications from person-to-person: the use of the human network model. Also the use of working papers or conference papers contributed to a relative efficient circulation of scientific information at low cost.

Electronic publications also offer the opportunity of experimentation to get away from a number of limitations which are intrinsic to paper publishing.

There are number of advantages. The most important is currency advantage. Then there is the relatively low cost of producing an electronic journal in comparison to a traditional journal. It has no printing cost and no distribution cost per copy put in circulation.

An important feature is the quality of 'gate-keeping' a security control to ensure the quality and going building on the status 'branding' of a journal. This is a difficult issue for electronic journal to tackle and also in relation to the aspect of asserting integrity assurance. How are electronic materials going to be archived for posterity ?

The peer reviewed status of important paper based scientific journals is still a threshold.

Many researchers with knowledge about E-journals still favor in the end the traditional journals with a good reputation. Because this is what still counts for researchers: the number of publications in the established journals.

Position researchers

Also the position of researchers as a group cooperating with commercial publishers is at stake. Researchers depended in the pre-Internet time heavily on the services of publishers and in return for that they demanded exclusive copyright from authors. This allows the publisher to control and charge for any subsequent use of the published material – even by the authors themselves. Scientists also depend on publishing for career advancement so they don't have an incentive to stop submitting articles. Afterwards the material can be reprinted or even electronically reused. This is a very nice deal for the publisher who pays none of the costs of originating his material and gets paid by the same libraries of institutions that are feeding the publishers.

E-print vs. P-print: a undecided case

E-print archives will be a model to follow from scientific researchers and responsible publishers. However a number of conditions will have to be met to be successful like I mentioned before.

The contents of an electronic publication needs to be careful organized to help the reader find the relevant information. Also the peer review of submissions is something which can't be missed. Last but not least the integrity needs to be assured.

It is interesting to know if e-journals have become accepted as a new tool for distributing scientific information. Despite library support the success of electronic publishing is not guaranteed. Electronic journals face the same difficulties to those of any new journal in getting established and attracting authors. It's also a matter of competing with similar printed publications. That doesn't mean that both types of journals can't be living together. In physics, the Los Alamos e-print repository – where preprints are submitted by scientists and are available for free – millions of transactions are made. These preprints have supplanted journals in distributing primary literature but didn't have a negative financial impact on the printed physic journals.

Two models of scientific publishing and grey literature

The picture which emerges is that there are two kinds of models of scientific publishing. The first model is the way in which grey literature is electronically produced. This is an idea-sharing model that benefits researchers in their exchange of knowledge and information and in public interest. Information is seen as a free good to benefit the quality of scientific publishing.²⁶ At the same time the idea-sharing model supports the notions about the scope and the objective of copyright as formulated within the US Constitution and the Recital of the WIPO Copyright Treaty

The commercial model could play a role in scientific publishing insofar a successful title from the online world would be transferred to a commercial publisher. A drawback would be the immediate substantial price increase. So the commercial model does the same in the online world as in the offline: just pay for what you read.

The question is what the future will be for grey literature in relation to publishing and commercial publishers. Do commercial publishers want to play an intermediary role also in the grey literature field in some way or another without turning the Internet into an electronic Fort Knox by means of taking over from authors their intellectual property rights ?

Their position seems difficult owing to themselves for digging in and defending their traditional publishing stronghold by lobbying for strengthen intellectual property rights. What they do is applying a kind of military strategy by not only protecting contents but also trying to control access to information. This is a new development but in my opinion from a strategically point of view a strange choice. This sound almost like the swan song, the farewell speech for traditional publishing. I would rather suggest if you can't beat the idea sharing people, join them !

It could very well be that the large commercial publishers are at end of their lifecyle and just don't know how to respond to the new challenges of the digital world.

The production of grey literature has set such an example and shows a method for a viable

way of publishing scientific information. One concern is still luminous. Large companies with more money could squeeze not-for-profit publishers out of the electronic journals market. This can only be solved by counterbalanced measures which would stimulate competitive practices. Perhaps the academic world in the US or Europe could vote for an agency ensuring principles of a competitive market but also protecting the interests of the scientific community in sharing information.

At this moment there are a number of not-for-profit setups to help universities and societies to publish at low cost like High Wire, an initiative by Stanford University Libraries and Academic Information Resources. It will take probably a number of years to reach a new equilibrium in the market before the not-for-profit approach is rewarding enough.

The Scholarly Publishing and Academic Resources Coalition (SPARC), set up in 1997 by the US Association of Research Libraries, is even more determined to engage the battle with commercial publishers by launching journals which are aimed at competing with equivalent expensive titles. Libraries who have teamed up with SPARC promised to buy each title²⁷

The Internet promises – or threatens perhaps – to change fundamentally what we mean by a book or a journal. And also the relationship among author, publisher and reader will change. Instead of a short term approach for implementing solutions to the opening of what looks like Pandora's digital box a more long term approach seems to be advisable to solve intellectual property issues.

As we had not much practical experience with the electronic media in general, it takes time to find a useful strategy how to operate legal intellectual property instruments in an electronic environment.

The objectives for a long term strategy are in my opinion twofold. Firstly we ought to adapt copyright law itself to new technological circumstances. Not by imposing old rules which are valid in an offline world and which are based on physical expression, but we should respond with new online rules in protecting copyright holders' interest in their creation which do justice to electronic works.

Secondly at the same time we should allow users to benefit from new technology. This adaptation is crucial to maintain the balance between the rights of authors and the public. Copyright law itself doesn't just apply to protection of authors but it has also a social value.

The working of the present intellectual property laws are based on territorial principles. So each country has its own set of 'rules'. So it is important to reach an international consensus on the scope of copyright. Intellectual property rights could be considered as kind of property rights reserved to the owner. Commercial publications rely on the private property mode, however not in a way this would obstruct the sharing of information as ideas. So the access to information would have to be better guaranteed for certain groups of users: students and researchers and teachers.

Uses of grey literature fall for a part in the public domain and might as well be given free away. Other parts of scientific publishing on the Internet deserve others forms of protection which conforms to the common property model. Common property model doesn't mean the absence of property. Authors of scientific works on the Internet can allow free use of their work for scientific communication by others, with reserving the right for republishing for commercial reasons.

This leads to the conclusion that the mentioned models of publishing can live together each

with his own set of rights. Only time will tell us which will survive in the end.

To end my paper I will tell you something about the situation concerning scientific publishing in the Netherlands. There the discussion goes on between scientists and universities regarding the possession of intellectual properties and the advent of the Internet.

The Universities claim that they are better equipped as an organization to make collective bargains and there therefore will be able to protect the interest of individual scientists. So universities expect that scientists will transfer their rights without any reward for this. Law by nature doesn't work in this way. Of course universities offer some advantages eg. scale, setting up consortia, negotiating licence agreement with publishers etc. But just giving away the rights by a scientist would only reduce the incentives for the scientist and harm the so called academic freedom. A better alternative would be to negotiate on basis of contract law, with properly rewards for the author being build in the contract.

The best option however would be if the scientists decide for themselves which mode of communication they wanted to follow and taking the influential role of grey literature into account – which blossoms outside the commercial sphere - the common property model wouldn't be a bad choice. Under common property the publication can be used by a particular group of persons, but is closed bu outsiders.

As scientific information is considered as kind of limited public good there have to be restrictions about how and how much each person may use it. Within these restrictions the information should be open to all members.

Thank you for willing and kind attention.

Autobiographical note:

drs. Cees de Blaaij is law librarian and head of the Documentation Center from the Institute for Information Law, University of Amsterdam, Netherlands.

Address:

University of Amsterdam
Institute for Information Law
Rokin 84
1012 KX Amsterdam
Netherlands
tel. 31-20-5253927
fax 31-20-5253033
email: blaaij@jur.uva.nl or cees@deblaaij.net

¹ The copying and abundant distribution of music recordings in a digitized, compressed format: MP3 throughout the Internet is one of the striking examples of the impact of a digital technology

² Noteworthy is the widening of the scope of copyright protection through the years since the Berne Convention came into existence.

³ Available at <<http://www.wto.org/wto/intellect/l-ipcon.htm>>

⁴ WIPO Diplomatic Conference on Certain Copyright and Neighbouring Rights Questions, Agreed Statements Concerning the WIPO Copyright Treaty (Dec. 23, 1996) <<http://www.wipo.org/eng/diplconf>>

⁵ Berne Convention for the Protection of Literary and Artistic Works, Paris Act 1971. WIPO, Geneva, 1989

⁶ Article 7: Scope of Right of Reproduction. Article 7(1): The exclusive right accorded to authors of literary and artistic works in Article 9(1) of the Berne Convention of authorising the reproduction of their works shall include direct and indirect reproduction of their works, whether permanent or temporary in any manner or form.

⁷ Article 10: Right of Communication. Without prejudice to the rights provided for in Articles 11(1)(ii), 11bis(1)(i), 11ter(1)(ii), 14(1)(i) and 14bis(1) of the Berne Convention, authors of literary and artistic works shall enjoy the exclusive right of authorizing any communication to the public of their works, including the making available to the public of their works, by wire or wireless means, in such a way that members of the public may access these works from a place and at a time individually chosen by them.

⁸ Information Infrastructure Task Force, Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights 7 (1995) [also White Paper]

⁹ EC Green paper 'Copyright and Related Rights in the Information Society', COM (95) 382 final, of 19 July 1995

¹⁰ Proposal for a Directive on Harmonization of certain aspects of Copyright and Related Rights in the Information Society, COM (97) 628 final

¹¹ For maximum security the U.S. Bullion Depository, a solid square bomb-proof structure with mechanical protective devices, was built there in 1936 to hold the bulk of the US nation's gold.

¹² Berne Convention: Article 9 [Right of Reproduction: 1. Generally; 2. Possible exceptions; 3. Sound and visual recordings] (1) Authors of literary and artistic works protected by this Convention shall have the exclusive right of authorizing the reproduction of these works, in any manner or form. (2) It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author. (3) Any sound or visual recording shall be considered as a reproduction for the purposes of this Convention.

¹³ Pub. L. No. 105-304, 112 Stat. 2860 (1998) (adding §§ 512 and 1201-05 to the Copyright Act of 1976)

¹⁴ See 17 U.S.C § 1201(a)(1)(A)

¹⁵ Id. § 1201(a)(1)(C).

¹⁶ Sony Corp. v Universal City Studios, Inc., 464 U.S. 417, 428 (1984)

¹⁷ 'The enactment of copyright legislation by Congress under the terms of the Constitution is not based upon ant natural right that the suthor has in his writings,...but upon the ground that the welfare of the public will be served and progress of science and useful arts will be promoted by securing to authors for limited periods the exclusive rights to their writings' (H.R. Rep. No. 60-22222, at 7 (1909)).

¹⁸ Vanna White vs Samsung Electronics America Inc, 989 F.2d 1512, 1513, 26 USPQ 1362 (Kozinski, J. dissenting).

¹⁹ John Perry Barlow is a retired cattle rancher, lyricist for the Grateful Dead, co-founder of the Electronic Frontier Foundation and professor at The Berkman Center for Internet & Society at Harvard Law School

²⁰. EFF, the Electronic Frontier Foundation, is a non-profit organization working in the public interest to protect civil liberties, including privacy and freedom of expression, in the arena of computers and the Internet. EFF was founded in 1990.

Their website is <<http://www.eff.org>>

²¹. Barlow, J.P., 'Selling Wine without Bottles. The Economy of Mind on the Global Net, in: P. Bernt Hugenholtz (ed.), *The Future of Copyright in a Digital Environment*, Den Haag : Kluwer Law International, p. 169,170.

²². Id. p 174

²³. Id. p.174

²⁴. As a democracy becomes more online and citizens are stimulated by governments to participate in the information society they will get more involved in the political process. From that perspective it is considered essential that government information is being put into the public domain and people will have access to relevant information. National parliaments stimulate or have been stimulating this process by introducing supportive legislation like the Freedom of Information Act in the US and the policy guidelines from the European Commission for the establishment of the Information Society project. In this way a part of grey literature produced from all kinds government agencies is in the public domain free from copyrights and very often electronically available on the Internet.

²⁵. Farace, Dominique J., *Rise of the Phoenix: A Review of New Forms and Exploitations of Grey Literature*, in: *Publishing Research Quarterly*, vol. 13, nr. 2 (1997)

²⁶. John Perry Barlow defended this model in his famous article J.P. Barlow, 'The Economy of Ideas. A framework for rethinking patents and copyrights in the Digital Age, in: *Wired*, 2.03, 1995

²⁷. *Nature* 393, 719; 1998

The Application of Electronic Copyright Management Systems (ECMS) to Grey Literature

Dave Davis, Project Manager, Copyright Clearance Center

Abstract:

Electronic Copyright Management Systems, such as those provided by Copyright Clearance Center (CCC), enable maximum dissemination and impact for copyrighted content (including grey literature), while still returning aggregated use data to rightsholders (authors, publishers or others) as well as a range of royalty options, from 'zero' to "whatever the market will bear." New and unanticipated uses of grey literature, such as republication into a collective work (essays collected in books), or licensing for digital use on corporate intranets, or transactional licensing for distance education or other academic use on digital networks (such as professors making their class notes available for use by professors at other schools), are often most efficiently and effectively managed by a third party. A third party aggregator, such as CCC, matches potential users with a repertory of permissions (rights to use different works in various ways) in a web-enabled, customizable environment. Several examples are provided, with questions for further research included.

In 1788, referring to the proposed constitutional clause for copyright and patents in Federalist #43, James Madison wrote "The public good fully coincides in both cases with the claims of individuals." Whether this idyllic description fits any specific copyright regime is a matter beyond the scope of this paper, but that copyright law and systems based on that law generally are intended to benefit both the public and private concerns seems indisputable. The familiar twin clauses of Section 8 of Article I of the U.S. Constitution echo Madison's understanding of this rare 'full coincidence' of the public and private goods: "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."

"To promote the Progress of Science and useful Arts,
by securing for limited Times to Authors and Inventors
the exclusive Right to their respective Writings and Discoveries."

(U.S. Constitution, Article 1 Section 8)

In this paper, in working through the example of Grey Literature reuse (republishing and redistribution to new publics), I will describe how Electronic Copyright Management Systems (ECMS) can be used to facilitate these noble and 'coincident' goals, both in 'promoting the progress of science' as well as 'securing' rights for private rightsholders.

Systems for managing rights and permissions, in the context of secondary publishing, have been around for many years. Both Charles Dickens and Mark Twain, for instance, successfully increased their fame by recasting material originally written for magazines into full-length books. More recently authors, literary agents and publishing organizations have had to develop routine procedures for answering requests for the re-use of materials that they control. In our new web-enabled, multimedia society the complexity of these requests, plus the potential for ever-increasing volume, implies an increasing need for more efficient systems to handle such requests. Rule-based automation facilitates this process, while honoring the constraints and requirements of both creators and licensees. ECMS were functionally described in extensive detail in a December, 1998 World Intellectual Property Organization (WIPO) paper, "Electronic Rights Management and Digital Identifier Systems" by Daniel Gervais (of CCC). Contemporary ECMS, such as those offered by Copyright Clearance Center (CCC) are able to handle the complex rights involved in licensing copyrighted content (text, audio, video, photographs, animation, streaming media, etc.). On the other hand, ECMS must also be able to meet rising user expectations that are now coming to include rapid, granular search & retrieval, bundled permissions, financial transactions, and content handling. Increasingly sophisticated, impatient users will require "Soup-to-Nuts" solutions, not handoffs and delays.

Although a technology-neutral, standards-based approach represents the path of wisdom to this goal, any complete market solution will tend to include at least these several technical elements, and perhaps others:

- Works level (Granular) Searching
- Standardized Metadata Tagging
- Persistent Resource Locator
- Maintenance of Rights Management Information
- Digital Content
- Real Time Access to Usage Data

Users will expect to search and locate at the article level or below. Downstream distributions (by users, as 'republishers' to new users) are possible, but not guaranteed. Standardized metadata tagging, as well as the use of persistent resource locators, enables potential users to find the material reliably, time after time, despite transfers of ownership or changes in distributors. Web-oriented users presume that the content they seek begins its existence in digital form, and continues through its lifespan as bits, not as ink on paper or celluloid.

Meeting these presumptions is beneficial to both users and rightsholders. A secure container option for the material may be utilized to assure rightsholders that this is not the last sale or license for their material- this is a distribution, not an "abandonment of rights". This option is likely to be more appropriate for materials of high commercial value, or for documents of a confidential or sensitive nature. The last element, real time access to usage data, is also of primary value to rightsholders who need to know how well their materials are doing, both as a commercial property, and as an offering in the marketplace of ideas.

These technical elements are only part of the picture: human interaction, or just plain old great customer service, is also a nearly universal expectation of users and rightsholders in functioning copyright management systems. What CCC is finding, as a first-mover in this market space, is that technology *alone* is decidedly not enough - it is necessary but not sufficient, as my instructors in scholastic philosophy would say. Incorporation of this point, that high-touch customer service plus functional user-friendliness embedded within the technology plus recognition (based on experience) that the real world is not entirely susceptible of automation, is what distinguishes real world ECMS solutions from 'pretty on paper' business models that have not yet stood up to the acid tests of meeting a bottom line and meeting customer expectations, day after day, scalably and reliably.

As a context for the prospect of new uses for copyrighted material, in a recent report prepared for the U.S. Copyright Office, "Project Looking Forward, Sketching the Future of Copyright in a Networked World" (May 1998), Trotter Hardy wrote, "an appreciation for possible future market effects of a new use technology is essential to good copyright decision making in the present...." Similarly, aggregated information about the use and placement of a work may, in some cases, be more valuable than royalties! As the papers from previous GreyNet conferences demonstrate, a document within Grey Literature most commonly begins its life as a report or other publication intended to fulfill the requirements of a relatively small audience. While a significant portion of Grey Literature may be in the public domain, many other Grey Literature works are copyrighted and theoretically available for commercial exploitation. Or, if not offered for sale in hope of profit, then they are simply made available for redistribution to a wider audience, for low- or no- fees. An ECMS enables these redistributions to occur without unnecessary delays or loss of control.

While grey literature, almost by definition, often begins existence in a non-commercial context, multiple re-uses are possible and indeed often occur in the real world. Grey Literature papers are often republished into books, developed into "new media" (audio, video, CD-ROM) or even incorporated into a "community of interest", in the sense made popular by Larry Downes, Chunka Mui in their recent book Unleashing the Killer App: Digital Strategies for Market Dominance (Harvard Business School Press, 1998) by being excerpted or

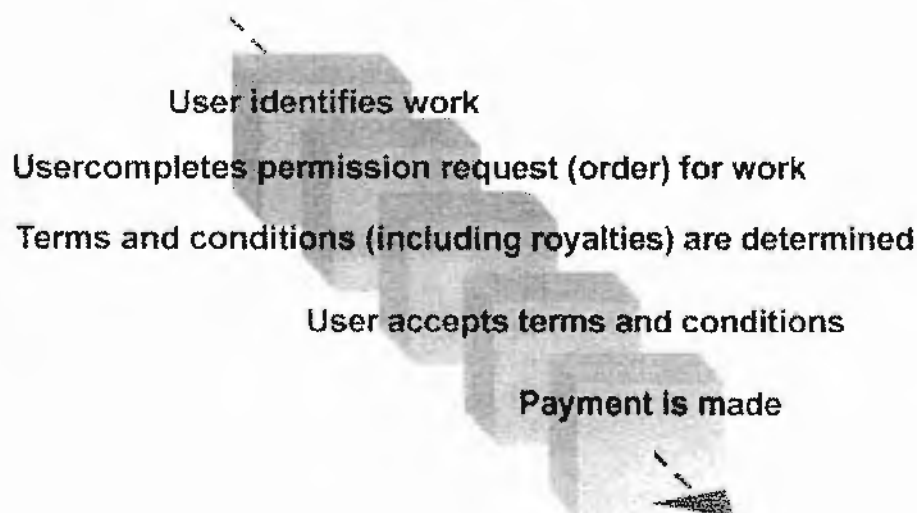
uploaded to a web-based discussion board. In that context, written works represent knowledge assets- they encapsulate findings, and, when of high quality, provide useful answers to real-world problems. They merit intellectual respect, and in the instance of Grey Literature, often deserve wider distribution than their initial publication achieves.

What are the prerequisites of a fully functioning ECMS? There are two principal aspects to ECMS- many ballyhooed commercial solutions seem to grab one or another of these aspects, but not both.

For the simpler aspect, the economic transaction, the requirements are:

- a Known, Discrete Work;
- a Rightsholder with 'Clear Title' ;
- a User with actual Intent-To-Use
- a Mutually Acceptable Agreement (and, possibly)
- a Payment Mechanism

Re-use as a Simple Economic Transaction



The second aspect, the communications cycle, entails a more complex system:

- Out of an earlier Work, a new Work is created by a creator for an audience; then
- The Work is made available (passively) or distributed (actively) to the audience; then

- The Work is received by its first audience- read, critiqued, acted on or disregarded, perhaps quoted or cited, or affirmed or rebutted; then
- The Work may be repackaged, translated, digitized or uploaded, excerpted or used in a new report, updated and revised, deeply annotated and discussed via hypertext; then, optimistically
- This now classic Work may become the basis of a new Work (Return to top of cycle and begin again)

Re-use as a Communications Cycle



This communications cycle is familiar to researchers and authors publishing in the sciences and social sciences, especially those utilizing peer review. From that perspective, any economic transactions that occur may appear of secondary importance, if they appear to bear any importance at all.

Copyright Clearance Center (CCC) is the largest, and the only not-for-profit, licensor of copyright reproduction and distribution licensing/permissions services in the U.S. CCC was formed in 1978 to facilitate compliance with U.S. copyright law. CCC provides licensing systems for the reproduction and distribution of copyrighted materials throughout the world. As part of our mission, CCC continually develops systems that utilize technology to support copyright compliance. Additionally, CCC's newly developed rights management technology allows us to respond quickly to opportunities on behalf of our rightsholders and users, and to provide a wide range of licensing programs tailored to various types of media.

Currently CCC provides all the elements of an ECMS in the sense discussed above through a suite of services:

- Direct Author Registration
- Various User Licenses: Transactional Reporting Service (TRS), Annual Authorizations Service (AAS), Academic Licensing Service (ALS)
- Republication Licensing Service (RLS)
- Plus others as well

Republication Licensing Service (RLS), in particular, fits well in this analysis, both from the perspective of *users* and from the perspective of *rightsholders*. The new RLS from Copyright Clearance Center is a fully automated, online service for licensing the rights to republish copyrighted works in both print and electronic media. In order to make efficient and effective use of previously published, copyrighted material, *users* have the ability to place their requests quickly (via web), to have the permission request either granted on the fly or, if necessary, automatically forwarded to the rightsholder. *Rightsholders*, on the other hand, gain the ability to have incoming requests made as standardized as possible, made for materials that they in fact control, and processed all the information necessary to making a permission decision. And, if royalties are part of the equation, they will be assessed accurately and according to the rightsholders' specifications. *Both* rightsholders and users benefit from knowing, in the aggregate, *what* material is used (a criterion of value), *who* is using it (a marketplace datum), and *where* (or to what new audience) it will be redistributed. CCC's existing services are capable of facilitating this complex process at a highly granular level (i.e., individual articles or photographs, etc.), and new service enhancements are planned to continue meeting these emergent user and rightsholder needs.

RLS also accelerates certain other user and rightsholder requirements for an ECMS. In the case of users, RLS obviates the need for time-consuming sleuthwork sunk into locating an obscure copyright holder. As an additional time-saver, in many cases the permissions will be granted instantaneously. Otherwise, RLS automatically contacts the rightsholder on the user's behalf, then reports the results to the requester online. Perhaps most importantly to users, use of the authorizations provided by RLS becomes a key component in their strategy of minimizing any risk of noncompliance. The additional requirements of rightsholders are: the option to decide which works to authorize for republication; what royalty fees to charge based on the type of use; whether to grant permissions on a pre-approved or case-by-case basis; and which formats (e.g., print, e-mail, Internet, Intranet, CD-ROM) to authorize republication within (based on the scope of rights held by the rightsholder). Finally, CCC distribution reporting aggregates the usage data and provides indications of the republished material's impact upon its new audiences.

Stepping away from the marketplace or service-level view, what are the ECMS functions described here?

- Aggregating users into a comprehensible marketplace
- Aggregating rightsholders and their works into a catalog of needed permissions
- Aggregating user data, works data, and rightsholder data into informative reports
- Processing financial transactions quickly and seamlessly
- Managing and maintaining rights knowledge accurately, and at the necessary degree of granularity

The complexity of administering this web of relationships is what renders rights management a field not for those faint of heart or new to the game. However, this overview demonstrates that CCC's existing services speed up and normalize *both* the transactional aspect – doing the deal, if you will – and the communications aspect. Additional services under development will do so with even greater efficiency and with wider scope.

Although CCC's services are online and operational *now* for the re-purposing of Grey Literature, several topics implicit in the discussion above remain as open areas for further research:

- To what degree is grey literature re-used in commercial publishing? What are the trends?
- What is the impact factor of grey literature, as measured by re-use? Does it differ from impact factor as measured by citation?
- What unexploited market opportunities exist for custom publishing of grey literature materials?
- Is standardized subject access (on a granular level) a prerequisite of increasing reuse?

Publishing and grey literature re-use are often analyzed transactionally. This works well from the perspective of economics, in Madison's sense of 'private good', and as e-commerce continues to advance we should expect to see more of this understanding applied. However, viewed as communication cycle, a different picture emerges– one of a feedback loop that continually increases the public good through its iterative turnings. It is the challenge of any ECMS, and of CCC's services in particular, to honor and facilitate both terms of this complex equation.

Finally, while we may not be able to agree with Mark Twain, who claimed that only one thing was impossible to God, to find sense in any copyright regime in existence, we can hold onto a vision where copyright systems enable and

facilitate both economic transactions **and** scholarly and professional communication, for Grey Literature and all others.

Additional Resources on ECMS

"Copyright Collectives and Libraries" (Interview) Copyright & New Media Law Newsletter (Spring, 1998).

"Copyright in a Digital Age: Practical Guidance for Information Professionals."
Robert S. Weiner, CCC Online, May 1997.

"Copyright Protection for Electronic Publishing Over Computer Networks". A.K Choudhury, N.F. Maxemchuk, S. Paul, and H.G. Schulzrinne. IEEE Network, p 12-20, May-June 1995.

"The Economics of Print versus Electronic Publishing: An Illustration from Mathematics"
Michael Rappa (draft) December 1998. URL:
<http://www.technika.org/copyright/model.html> ; see also the accompanying graphic
<http://www.technika.org/copyright/model.pdf>

"Electronic Rights Management and Digital Identifier Systems," by Daniel J. Gervais
Journal of Electronic Publishing, University of Michigan Press (March, 1999)
URL: <http://www.press.umich.edu/jep/04-03/gervais.html> .

"ERCOMS (1998) Market Survey of Existing Electronic Copyright
Management Systems (ECMS) and Projects."
URL: <http://www.iieir.dmu.ac.uk/Projects/ERCOMS/ercomsm.html>

Information Sources in Grey Literature . Peter Auger . Publisher: Bowker-Saur. 4th edition 1998.

"The invisible hand of peer review." Harnad, S. Nature [online] (Nov. 1998)
URL: <http://helix.nature.com/webmatters/invisible/invisible.html> .

"A new generation of Grey Literature: the impact of advanced information technologies."
G. A. Cotter, B. C. Carroll. Paper presented at GL '93 (December, 1993).

"PURLs: Persistent Uniform Resource Locators" Stuart Weibel, Erik Jul and Keith Shafer.
URL: <http://purl.oclc.org/oclc/purl/summary> .

"Rise of the Phoenix: A review of new forms and exploitations of Grey
Literature." Dominic J. Farace Publishing Research Quarterly. Vol. 13, #. 2, 1997.

"Survey on Electronic Rights Management" (1998 -) International Federation of Reproduction
Rights Organizations (IFRRO), URL: <http://www.ifrro.org/committees/surveyintro.html>. See also
IFFRO's Collective Management of Digital Rights (October, 1996)
<http://www.ifrro.org/papers/pp-digi.html>

Copyright: Black and white or just making you see red?

**Graham P Cornish,
Copyright Officer, British Library and
Director of the IFLA Programmed for Universal Availability of Publications**

We live in an era in which rights are a matter of constant discussion and debate. But rights are far from uniform in their character and it is important to place the right to copy – copyright – in its proper context in order to understand some of the issues surrounding grey literature in its broadest sense.

We acquire rights in several different ways. There are what most people regard as the basic human rights – food and water, shelter, and the other basic necessities of life. We acquire these by virtue of our humanity and they require no action from us, nor should they need any legislation to enforce them. Other rights are bestowed upon us by the law of the land such as rights for compensation when we suffer inquiry or redundancy. These rights are ours whether we want them or not or whether we need them or not. We need to exercise them only when the circumstances are appropriate and we may choose not to do so. Then there are the rights that we acquire deliberately such as those that go with a piece of land or a trademark. Of course, occasionally we do buy something and acquire rights in the process we never knew existed. The case of the Birds' Eye Company buying a farm in Lincolnshire (England) and finding they also acquired the right to appoint the vicar of the local church is a quaint example!!

Copyright is a right which manifests several of these different characteristics. Whilst nobody would claim it was as fundamental as food and shelter, it still underpins the basic concept that what we create we ought to be able to control. This is because what we create is a projection of "self" in a very vulnerable and obvious way and it is usually considered a basic right that we can and should be able to protect ourselves. For those countries which have signed the Berne Convention¹ the rights conferred by copyright on the author are automatic. They are defined by the legislation of the country concerned and the author, as first owner of the rights, has no say over what rights are acquired in this way. Copyright is conferred by the state in such forms and within whatever legal context the legislators consider appropriate. Once a work has been created the author acquires certain rights (which may vary in content from one country to another) but which are absolute. Many authors never realize they have these rights, especially those who do not create works for gain, either financial or academic. Similarly, many users of copyright material never consider that any rights exist in them at all and those working in the field of copyright can quote example upon example of bizarre and strange ideas about copyright and the ways that one person wishes to use the intellectual property of another.

It is in this context of rights acquired whether they are wanted or not that I would like to discuss copyright and grey literature for I believe it is this passive acquisition of rights that can cause so many people to "see red" about copyright when discussing grey literature in its widest interpretation. There is no point in trying to define what grey literature really is. This is a matter of frequent debate and the general definition of something which has not been commercially

published for general distribution through the normal retail channels will suffice for the purpose of this paper.

Because copyright is automatically bestowed on the creator of an original work, that creator has no say in what rights are acquired or how s/he may wish to exercise them initially. Essentially copyright was invented to protect two distinct elements of the creative process: firstly, the economic value of whatever was created; second, the so-called "moral" rights of the creator, namely the paternity and integrity of the work created. In broad terms this means the right to have authorship acknowledged and the essential content of the work to be left in tact and not interfered with. The distinct nature of these two bundles of rights is a further complicating issue when looking at grey literature. This complication comes about because of the different ownership of these rights, depending on various legal and political circumstances, the complexities of which need not detain us here.

Authorship is, of course, central to the whole idea of copyright because, until something is created there can be no copyright. One of the distinctive aspects of copyright is that it subsists, rather than exists. In other words it is always necessary to create an object (in which other rights of ownership, for example, may reside) before copyright can subsist. [Perversely, copyright CAN subsist in something which no longer exists although it did once in the past. The classic example is a painting by Graham Sutherland of Sir Winston Churchill which Lady Churchill destroyed as soon as he died - but the copyright in the painting still subsists because there is a photographic record of it]. Most authors of grey literature have little or no interest in the economic exploitation of the publication they produce. In the first place, any rights in it are probably owned by their employer as discussed in a moment and secondly, they have no personal interest in the value of the actual document except in one crucial area. Although the so-called moral rights for authors may not be available to employees in some legislative regions, most authors of grey literature will have a strong interest in being acknowledged as the author of all or part of the document. Reputation for researchers in all fields, whether in the natural sciences, social sciences or the humanities is central to developing an acknowledged role in the specialized field in which a particular researcher works. Failure to be acknowledged as the author of a work, however esoteric or trivial it may seem as a publication, can result in serious loss of prestige in professional and academic circles with the resultant loss of status and possible career development. Therefore authors of grey literature actually have an interest in their work which is quite different from that which some employers or owners of economic copyright will have.

The other interest that authors have is the integrity of their text or other material. Statistical or chemical data can easily be manipulated to demonstrate quite different findings from those of the original researcher and the removal of one or two key phrases or even the word "not" will change the meaning of the text for ever. Therefore the integrity of the text is probably at least as important to researchers in all fields as acknowledgement of authorship itself.

So, now let us look at ownership, as this will determine attitudes to the economic elements of copyright. In most jurisdictions copyright in a work is owned by the employer if the work is created as part of employment, otherwise the copyright belongs to the person who created the work. So the copyright in a detailed report on a new drug prepared by a scientist working for a major pharmaceutical company will be owned by that company. But the company is far more

likely to be interested in the content of the research, which is not protected by copyright, than the copyright in the text itself which IS copyright. Once they have obtained all the information they need from the scientist and the report they are less likely to be interested in the text of the report itself. Having been cleared by the appropriate departments in the company from a confidential point of view, the company will be interested only in so far as no other company tries to copy the processes or research described in the report. The fate of the report itself will usually be of little or no interest. It may be released through a number of channels such as NTIS or the British Library or simply put on the company website where it can be freely downloaded, printed off, copied, stored or networked to any number of interested parties. The company may well not mind any of this happening although it infringes several distinct rights enshrined in copyright including copying, distributing to the public and perhaps broadcasting². This is in stark contrast to the commercial publisher who has a strong economic interest in the distribution of any work they publish. This may include hardback or paperback rights as well as geographical areas of distribution and the possibility of turning the book into a film or TV series. For academic publishers they will have an interest in mounting a scientific journal or report on CD-ROM to be sold for profit or at least putting it on a website to which access is limited to those who agree to pay. Very often the producer of grey literature will be grateful that someone copies or distributes the work as it actually saves the original company money in terms of staff time, consumable and warehousing. Where a work is made available electronically then, again, most producer of grey literature would be unwilling to invest large sums of money or technical expertise in constructing mechanisms to inhibit copying, downloading or distribution when these activities are actually the real reason the work was released in the first place.

However cavalier companies and institutions may be about the release and distribution of their reports, never the less they too may be sensitive about being acknowledged as the source of the information and may want their contribution to the increase of knowledge properly noted in any re-distribution of literature they produce.

We have therefore two sets of rights: moral rights which both authors and institutions will almost certainly wish to defend and economic rights which may be of little or no interest to those who actually own them. If they really saw any economic benefit in releasing material presumably they would have opted for the commercial publishing route anyway. This, in turn, leads to another approach to protecting copyright. Owners may not wish to go to the expense or bother of publishing material commercially but, equally, they may wish to prevent other from doing so as well. It is one thing to allow someone to download or copy a report for research or academic use: it is quite another to allow them to do the same actions for the purpose of commercial republication, either as a "stand-alone" document, or, more likely, to be integrated into a larger work either as a chapter for a book or part of an anthology on a specific topic.

But there are other concertinos in the field of copyright as well. Particularly important is the right of distribution. Owners of copyright have the right to control how their works are distributed or published. Although many owners of rights in grey literature may not have an economic interest in distribution, they may have another agenda to follow in terms of controlling *who* has ready access to the material and also *in what form* that access shall be. This may be because an institution wishes to discourage some other organizations or groups from gaining ready access to the material

or it may simply be that the owner wishes to know who wants to have access to give a picture of those working in a specific field of research. This links into the proposed³ "Communication to the Public Right" which is likely to hit the statute book in many countries in the next five years. This right will give the owner the exclusive right to make material available electronically, such as on a website, whether or not that material is actually consulted. In the normal document supply world the algorithm is used that 80% of demand is met by 20% of the collection and if this is applied to the World Wide Web there are an awful lot of websites or bits of websites that are never visited and this is almost certainly true. However, just to put something on a Website will be an infringement of the owner's rights even though it may never be read by anyone. At the moment there are arguments about whether or not putting something a website actually constitutes publication⁴ but this new right will take care of this legal nicety.

The creator and owner of grey literature can therefore demonstrate a whole range of rights which are acquired automatically upon creation of the work concerned. Neither the author nor the institution sponsoring the report has any control over what rights they acquire - they are automatic and given by law. Yet they may not wish to exercise any or all of them or may wish to do so only in very limited ways. The other side of this coin is that the user may understand that the owner has this bundle of rights and, equally, is unable to determine which of them the owner wishes to enforce, if any. Unlike physical property there is no ready way of signaling this. After all, if I own a plot of land I can put up a notice which says "Anyone may walk across this land" or, as a cricket field in south London in the UK has "You may walk your dog on this land but do not let it go on the grass." the variant meanings of the word "go" in this context not being explained. But for material which is copyright the user has no guidance. S/he may make an informed guess as to what the copyright owner might allow or consider acceptable but there is no guarantee. Frequently users consult specialists in copyright, only to be told what the law says but not what is the probable attitude of the owner. To be fair, who is prepared to guess on behalf of someone else in such a litigious area of law? They are then left to guess what might be possible without being taken to court. But nobody can be totally sure when guessing, otherwise it would not be a guess. Therefore the potential user is faced with the decision either to stick with the letter of the law in almost certainty that the owner will not take such a restricted view, or go ahead and do whatever they want to do with no assurance they will not bring down the full weight of the judiciary upon their heads.

Clearly this is an unsatisfactory situation when owners do not want to defend many of the rights given them by law and users usually want only to make use of a work in a sensible and realistic manner. What then is the solution?

It seems to me that some kind of internationally agreed code of practice on copyright for grey literature should be adopted. It would acknowledge the needs of users but pay due respect to the legal rights of owners as there will always be those who feel the need to enforce their full suite of rights. Essentially we end up with a table of needs and rights which looks something like the following. Naturally there is plenty of scope for refinement and additional situations to be included but this might provide the starting point for further discussion and implementation.

Action\Right	Code	Personal use	Internal distribution	Non-commercial publishing	Commercial publishing	Website use
Code		1	2	3	4	5
Always ask	A					
Acknowledge author	B					
Acknowledge source	C					
Paper copying	D					
Electronic copying	E					
Paper distribution	F					
Electronic distribution	G					
No restrictions	H					

By devising a set of agreed symbols using this sort of matrix would enable producers of grey literature to put a code on each document which could be understood internationally regardless of language, legal tradition or any other local circumstances. The code would act as a sort of proactive permission removing the need for users to seek unnecessary permissions and owners from having to handle requests in which they have no interest and which are nothing but a bureaucratic burden.

For example, using the highly simplified table above, an owner who did not wish to prevent electronic scanning of a work for personal use would code it E1; a document coded B5 could be used on another person's website provided the author was always acknowledged and so on.

Users would need to be assured that such a code was accepted by owners and used only by those who owned the rights that they were effectively licensing to others. But such issues could be sorted out if there was an international will to do so. After all, the initial reaction to various codes for serial titles were greeted with scepticism but eventually found a permanent place in the world of library science, cataloguing and publishing.

Naturally a lot of work would need to be done on refining this matrix which is just presented as an idea and the different possibilities would be much larger and the restrictions placed would need to be refined but I believe this offers a basic way forward for the management of copyright in grey literature in the future. It is offered in the hope that copyright will no longer be seen as "black and white" ("yes you can" or "no you can't" depending on the attitude of the person consulted) nor will it make either owners or users see red any more.

¹ International Convention for the Protection of Literary and Artistic Works (Berne 1886 and subsequently revised) Article 5(2)

² The question of whether providing access to a work through a website is broadcasting hinges partly on whether a website is a broadcast or a cable programme service. This was a major unresolved issue in the *Shetland Times* and *Shetland News* case. Innes, John & McMillan, Fraser (1997) "The *Shetland Times*" Internet case. *Copyright World*, no 76, pp26-28

³ Proposed by the WIPO (World Intellectual Property Organization) treaty in December 1996 and currently part of the proposed European Union reforms of copyright legislation.

⁴ The case of Nottinghamshire County Council in the UK trying to stop a firm of consultants from putting a report on child abuse on the web because it constituted publication is just such a case. CORNISH, Graham P. Copyright and the Internet. In: *Theatre and the Internet: proceedings of a conference held at the University of Glasgow....May 1997*. pp.24-28.

Issues in Distribution of Grey Literature: Experience of the Japan Documentation Center (JDC)

**Ichiko T. Morita
Japan Documentation Center**

The focus of this paper is to present some issues in grey literature, its dissemination, its place in the collection, grey literature in electronic formats, and difficulty in dealing with languages of publication. Those issues are presented from the past five years of the author's experiences in establishing and developing the Japan Documentation Center. The Center collects, processes and disseminates Japanese policy related grey literature in the areas of politics, economics, business and industry, the environment, and national defense. The discussion, research results, and suggestions follow each.

The first issue is on grey literature and its users. Users of such information like Japanese policy related literature are not necessarily scholars and specialists in Japanese studies. They encompass lawyers, economists, legislators, scientists and many people of other disciplines who must deal with Japan in their own special field. Some are not even regular library users. To disseminate the information on the functions of the Center is a great challenge because one cannot pin-point target groups such as the Association for Asian Studies and the Council on East Asian Libraries.

The second issue is the place of grey literature in the collection development policy in relation to other Japanese materials in a collection. In the recent budgetary crunch in many libraries, the criterion for collection development are reviewed and tightened in general to its core materials of research value in the long-term perspectives of the library, whereas the value of grey literature is for its timeliness and urgent need for practical purposes but not necessarily for its historical value. How the library managers justify the efforts and costs needed for collecting grey literature is interesting.

The third is the development of electronic information distribution and coping with it. Besides the copyright issue which the author does not discuss here, establishing the good practice in handling increasing number of electronically distributed grey literature in relation to paper copies should be thought of.

The fourth is the language of the grey literature. This issue relates to all the above issues; users, collection policies, and electronic distribution of grey literature. In the case of the JDC in which the major portion of literature is in Japanese, how to reduce the difficulties of users dealing with Japanese is another serious matter for some users.

The author discusses each issue area further in detail, with analysis in some cases, and brings some findings or ideas.

Introduction

In conceptualization of the idealistic mapping of the dissemination of grey literature, one quickly realizes several issue areas which must be sorted out in order to seek out ultimate solutions. The focus of this paper is to discuss some serious issues in handling Japanese grey literature, based on the author's past 5 years' experiences in the Japan Documentation Center (JDC). The Center was created as a part of the Asian Division of the United States Library of Congress (LC) and began its function in March 1994. It is established with a special mission to collect the most current grey literature on Japan from primary sources in Japan, concentrating on the policy related information covering the areas of politics, economics, business and industry, society, the environment and national defense. Its functions are to process the collected information quickly, create bibliographical access, and disseminate the information in the most timely and effective manner. Thus, the issues the Center faces may not necessarily be identical with those of more clearly defined academic and research libraries. However, the author believes that some of those issues are of mutual concern and hopes that the JDC's experiences may offer insights to others who handle grey literature.

I believe it might be helpful to consider six specific issues that fall under three broad categories. Some issues are intrinsic to the nature of the grey literature. They would include:

- 1) Scope of users
- 2) The language of literature

Some are based on the situation where the grey literature collection is a part of a traditional library as a whole. They are as follows;

- 3) Shifting focus to address changing needs
- 4) Ephemerality of literature vs. historical value of research materials

Lastly, some relate to other issues which emanate from the issuing organizations' practice. In this category, the following is included:

- 5) Printed vs. electronically disseminated information
- 6) Need for archiving

There are numerous other major, as well as minor issues, a major one being copyright related matters. However, this paper will not deal with them, the underlying reason being that such matters are dealt with in a different arena and the discussion here will not be very productive. They will be touched upon as they pertain to my topics being discussed.

Scope of Users

Unless an agency is a long established one and well known for collection of grey literature, the process of reaching users and educating them on the nature and availability of the particular grey literature collection can be rather difficult. In case of JDC, which began from a zero base five years ago, developing a systematic approach to identifying users and

disseminating information about the Center has been a serious issue throughout its early period. The fact that there has been much expressed need for Japanese policy information in the United States was a basis for establishing such a service. Moreover, providing for the urgent information needs of the U.S. Congress was the driving force behind establishing the Center at the Library of Congress. Once the JDC began its operations, however, it was extremely difficult to locate the diverse users who need the kinds of information which can be obtained only in form of grey literature.

Academic libraries and most research libraries can concentrate on serving their homogeneous constituents. In a Japanese studies library, for example, expected users are the Japanologists, including scholars, researchers, librarians and students. In case of the Library of Congress, users span a range of disciplines, and the JDC's services extends to everyone from Congress members to general public.

Characteristics of user groups, and issues

Among various categories of user groups, the academic researchers in Japanese studies were the only group the Center could target fairly effectively because they could be reached through already established networks. The JDC used two types of major networks: the Japan related Listserv's, each of which include Japan scholars, researchers, librarians, and students (though not limited to these groups); and professional associations, such as the Association for Asian Studies (AAS), the Council for East Asian Libraries (CEAL), and the Special Library Association. Messages are regularly posted on those Listserv's to inform their members of notable activities and new developments. The JDC staff make presentations and give demonstrations at those associations' meetings, and write articles on the Center's resources and services in various organizational publications. According to our informal studies, 39% of requests came from the academic community. The majority of these researchers focus their studies on Japan in such fields as political science, sociology, business and economics.

This means that the other 61% of users are non-academic and can be broadly considered under the general public category. We have found that these users are scattered in wide area, geographically covering the entire world and spanning numerous disciplines. In other words, there is no specific concentration. The population which needs and can make use of Japanese information, particularly in the Japanese language, has never been high, and there is no easy way to pinpoint who and where they are. The types of general public users of Japanese grey literature information include lawyers, technical experts, journalists, government administrators, exporters and business people, and other types of people who happen to deal with matters pertaining to Japan, among numerous other subjects. They would not likely know that such a specialized center like JDC exists, and for JDC, there is no systematic way to reach out to them. Some of them do not even think about libraries as an information source and some are not regular library users. Once the JDC home page was established in 1997 as part of the Library of Congress home page, it was much easier for users to learn of a specialized center like the JDC through their own web searching.

Another point which should be noted is that most of the users in the general public categories have been one-time users or extremely infrequent users, because they do not concentrate only on matters Japanese. Their needs are not constant. Therefore, it may seem that nurturing and encouraging such irregular users is wasteful, because they may never need the specific type of grey literature information again, and even if there is a next time, they might deal with a very different Japan related matter. A law firm is a good example. Once it has learned of the JDC's

services, it finds the Center very useful whenever it deals with a case with Japan. It may be arbitrary, however, as to when the firm takes on Japan related cases. As another example, an exporter appreciated the information we supplied very much because he needed to have information in order to begin exporting a certain item to Japan. Once that objective is accomplished, there is no further need for the JDC. Even in the Congress, the concerns with Japan are not continuous, though we regularly remind the Congressional members and their staffs of our services.

Outreach to users

Here we begin our endless outreach efforts. Even while the Center has built a strong grey literature collection, use of the information is a crucial measure of the Center's success. We have carried out the routine public relations steps, such as distributing brochures and newsletters and issuing announcements on Listserv's, and establishing a web presence. We routinely send printed literature to individual Congressional members and their staffs as well as to the Asia related committees. We maintain an international mailing list of think-tanks, academic and research institutions, organizations, agencies, and notable individuals, and include all that may have even a slight interest in contemporary Japan and public policy issues.

The JDC staff have presented papers on the Center, often accompanied by demonstrations, and have conducted separate workshops at local, regional, national, as well as international, conferences, meetings, and discussion sessions. To reach as many users and potential users as possible, we actively make professional linkages to organizations to be invited to present the JDC to their members at their meetings. In the past several years, these have included the U.S. Interagency Language Roundtable, Association of Japanese Business Studies, and the European Association of Japanese Resource Specialists.

The JDC has itself organized several very successful international symposia at the Library (1994, 1995, 1996), reaching out directly to a huge number of targeted and potential users through both printed and electronic measures. The symposia brought together experts to speak on a range of issues surrounding Japanese information and provided an excellent forum of exchange between providers and users of information, including grey literature. The JDC also hosted the Fifth JOHO Conference at the Library in 1997, in collaboration with the U.S. Department of Commerce's Office of Technology Policy. Publication of the symposia papers provides a further dissemination of information about the Center and on pertinent issues for those involved in research on Japan.

To summarize, it is clearly difficult to build a satisfactory number of so-called 'regular users' in a short time, in view of the reasons above. At the same time, those investing in and supporting the Center, who desire a quick and concrete outcome, look at certain measurements, such as numbers of users, numbers of documents delivered, and statistics on Congressional use. It has taken five years for the Center's existence to be recognized. We have found that constant, frequent, extremely patient and repetitious, reminders are necessary, and even so, usage patterns are not always consistent.

In addition to successfully making contact with users, there is a need to provide them orientation on how to effectively use the JDC collection and services. Many users are experienced and use the JDC resources either to locate specific information on well-focused topics or as a starting point to develop their research topics. We also direct them to other

resources as appropriate. On the other hand, we often receive questions beyond the scope of the JDC (e.g., historical, cultural), or inquiries from those who expect the Center staff to spoon-feed all the possible relevant information to them. The JDC home page describes the scope of the grey literature collection clearly explains how to request information and how to search the database. All the literature about the Center, printed and electronic, states that the reference and document delivery services are provided free of charge. We were therefore somewhat surprised that in a survey conducted in 1996 about the JDC, a number of people said the information is costly and that they could not afford to pay for the services. Such aspects of actual user behavior indicates to the JDC staff that continuous orientation is needed.

The Language of Literature

With the development of technology and subsequent capability of almost instant transmission of information in a global society, the language (e.g., Japanese, English) of the literature takes on dimensions. It is important to note that the nature of grey literature, which is often primarily aimed at a particular group of people, is that it is not necessarily issued in an international language, such as English. In the case of the types of documents the JDC handles, namely Japanese documents, these are mainly for Japan's policy makers and not particularly intended for a wider domestic or international distribution. More than 95% of the JDC's documents are in Japanese. As mentioned in the scope of users above, users of grey literature information at the Center are not limited to Japanologists who can read Japanese. In fact, over 60% do not read Japanese.

Thus the language abilities (more specifically, language limitations) of the users must be considered. Although the Center was established by the directive of the U.S. Congress which urgently needed Japanese policy information, we soon found out that there were very few Congressional members or staffs who can use documents in Japanese. In the Congressional Research Service of the Library of Congress where researchers investigate and prepare reports for the Congress on specific topics, there are several people in various disciplines who specialize in Japan. Even among them, not all can use JDC documents as they are. Most others need documents translated; some need them summarized; some need to wait for an English version, an availability of which is only arbitrary and certainly less timely.

Among JDC users, except for most of the academic researchers and a few federal government officers who specialize in Japan studies, the majority does not read any Japanese. However, it should be noted that some types of information are very useful for those who have some background on Japan but are not necessarily able to read Japanese. For example, the researcher who needs only statistics can successfully use them with only the translation of headings.

Minimizing language as an obstacle to access

As a partial solution, the Center provides an English abstract and assigns index terms for each document it acquires to enable users to easily identify potentially useful materials. The user, regardless of language abilities, can quickly grasp the content of a particular document and determine its usefulness from the bibliographic record before devoting time and effort to deal with the Japanese language document itself. This approach on the part of the Center requires special human resources.

The work of abstracting, namely going through a Japanese document, identifying the most

crucial aspects of its contents mentally, and writing an abstract in concise English requires quite an effort beyond routine abstracting work. Essentially, we do not have an option of finding experienced abstractors and to teach them the Japanese language. Instead, we have to search for people with Japanese and English language capabilities and teach abstracting. One soon realizes that it is much more difficult to create an abstract than to translate verbatim. It takes time and experience to become a good abstractor. Finding the appropriate personnel who can accomplish this work is not easily achieved. The work requires a native level of understanding more formal, written Japanese documents; training to extract the essence of information; and again native level of English to write clear abstracts. These steps must be done in a shortest possible time frame. For managing any grey literature collection, it should be kept in mind that this approach, while effective for the user, is nevertheless costly in terms of human resources and time.

Based on its experience, the JDC has found that greater ease of access to the content of the document collection, particularly one in a non-English language, has a direct impact on usage of the materials.

Users' language abilities shape their needs

At the time of the review of the Center in 1996 to 1997, we found conflicting user opinions: pros and cons of providing abstracts, or the quantity versus quality issue. One opinion is that rather than spending so much effort in creating abstracts, the Center should devote those resources towards acquisition of more documents. The other is that the abstracts are extremely valuable in locating and verifying how useful the information might be before the user begins to cope with Japanese language documents. It is obvious that the former comes from the group of Japanese scholars and librarians with facility for reading Japanese. The latter from the group of non-Japanese readers and non-specialists as far as subjects on Japan are concerned. However, even among the Japanese specialists, there are voices that point out that the abstracts are useful for quick identification of documents, given the general or vague nature of many titles of public policy materials (e.g., "Towards an Affluent Society"). Hence, it was decided by the Review Committee that we continue providing abstracts as a value added feature of the bibliographic database.

It is clear in the Center's experience that provision of English abstracts to Japanese documents improves access to the source information and considerably reduces the seriousness of the issue of language as an obstacle. Without the abstracts, there would be many who would not even try to approach to the information. This was attested to by several users who stated that they would have otherwise had to compromise with the secondary sources in English coming from English speaking areas (such as Southeast Asia) about Japan's policies, even though they realize that the Japanese source documents would provide the precise information on policy matters. As a way to access this body of policy information, these users found the JDC database with abstracts for each item in the collection highly valuable.

Under the present stage of development of automated machine translation of Japanese into English, the output, especially for literature in the social science fields, is yet inadequate. The Center's approach seems to be the next-best solution. The author strongly believes that providing English abstracts of Japanese documents is a notable value of the Center's database, regardless of many problematical factors, such as scarcity of qualified abstractors, cost of abstracting, the time it takes for abstracting, interfacing of abstracts into the database, and maintenance of the database.

The above two issues, namely matters emanating from the types of users (and their language abilities) and from the language of the collection, are based on the intrinsic nature of the grey literature. In addition, the following issues arise from the fact that the grey literature collection is a part of the traditional library as a whole.

Effect of Political Changes

In general terms, the following issues concern the subject matter of a particular grey literature collection and how it relates to changes in need for that literature along with broader societal changes. In the case of the Japan Documentation Center, it is obvious that the need for policy information, especially the kinds that the Center collects, is affected by the political situation as well as by the economy and other factors. To mention a few examples, when the U.S. government negotiated for various aspects of Japan's deregulation, the information on the topic was in great demand; and when the political focus of the U.S. government shifted from Japan's internal policies to Japan's economic policy towards the Asia region, the requests for the Japanese documents on private business activities, global environmental issues, economic assistance, and other related information increased. Avidly following its mission, the Center collected timely information focusing on the political, economic, and other topics in demand at a given point in time.

As a result, the Center's collection is not one of those that collects systematically on pre-determined topics in order to build a well-balanced collection for more historical and fundamental research. On the one hand, to maintain a relevant grey literature collection, the collection focus must shift as the needs shift. A possible short-coming of these circumstances, on the other hand, is the difficulty for the Center to pursue a certain topic and collect documents exhaustively. At the time of the Center's review, one reviewer pointed out that the Center lacks some reports in a series of report documents in a topic from the historical point of view. The Center tries as much as possible to collect all the related documents, but because of the nature of distribution of those documents, it is often impossible. For example, if policies on airport use come to be considered for revision, it is almost impossible to go back to the original policy documents and policy discussions, as well as subsequent revisions throughout the years.

The consequence of these two factors, collecting to meet changing needs and difficulties in exhaustive collecting due to the nature of how grey literature is distributed, must be understood as being both part of the strength and the limitations of a collection. It may be said that this is inevitable and stems from the nature and purpose of the collection, and therefore the grey literature collection must be considered within context the institution's whole collection.

Ephemerality of Literature

A common concept is to include the grey literature genre in so-called broader "Ephemeral literature" grouping because of its value being extremely time sensitive. Some information in grey literature is compiled and later incorporated into a more permanent form, such as annual volumes issued at a certain time of the year; some information becomes obsolete; while some others stay as they are. Providing timely information in a form of grey literature is highly

valued, but when evaluating it as a permanent collection of a library, quite different factors come into play. Under the current situation where many libraries are economically stringent, the retention and maintenance of collected grey literature simply come to be a matter of priority in terms of the library's broader collection development concerns. Grey literature must compete with other materials as to whether, or to what extent, efforts should be made to bibliographically control them as well as to expend costs for preservation measures.

Whether to absorb grey literature into the permanent library collection after fulfilling the role of providing timely information or to discard them is an issue that must be addressed. It is important to note that this is a library dependent decision. The policy of "once collected, never throw away" is no longer valid. Each library must make a wise decision, because the cost of retention is not nominal. In the Center's case, it has been determined that the document collection is to become a part of the Library of Congress' permanent collection.

Preservation of materials

Under this scenario, first the decision must be made as to how to preserve them. One of the characteristics of printed grey literature documents is that none are hard-bound, since generally speaking they are not meant to be kept and maintained. Many are issued in acidic papers and require some costly preservation measures. Furthermore, the spines of the original paper materials must be cut in order to be scanned. It costs much to render appropriate measure to each document and again, it has to compete with the rest of library materials which are of highly historical value for future research.

The Center fortunately has some funds of its own and is able to place the original printed documents in acid-free folders and in acid-free boxes and to organize them for permanent storage. However, the amount of funds is not large enough to de-acidify each document.

The possibility of microfilming can also be considered. In the case of the Center, the bulk of the documents has been digitized and the information can be easily printed out from the image database file. Therefore, it is justified to use the basic measures described above for retention of the paper documents. Furthermore, the image database itself can also serve as one of the preservation measures, although digitized images have not yet proven to be permanently secure.

Bibliographic control: the online catalog versus a special file

This brings up a second decision that must be made regarding access, i.e., bibliographical control. In the JDC's case, its bibliographic records are maintained in its own, separate database, and the structure of the record has been designed for the Center's particular needs. For example, the translated title appears first and is the one that is displayed in the listing of search results. Also, the Center has adopted the Legislative Indexing Vocabulary, developed by the Library's Congressional Research Service. If the materials are to become part of the permanent collection, the documents in printed format need to be processed, this time according to the Library's normal practice in order to be kept in a more permanent and regular book stacks. This means new catalog records must be created for each document and added to the Library's online catalog. For that purpose, ideally each record must be created following the current cataloging rules and contain various access points according to the pre-established headings. A degree of duplication of processing and cataloging may be unavoidable.

It should be noted that the reason why the Center painstakingly creates the bibliographic

records with index terms and English abstracts is that the access to collection is to be more orientated toward specific users of the type of information as well as towards a general duration of time in which it is needed. The approach is specifically tailored for the needs of the users, the majority of whom are not regular library catalog users. The Center's database which contains bibliographic records of those documents are now on the Internet as a part of the Library of Congress Web and is searchable. However, it may not be considered as a part of the permanent library collection. It is considered a special file.

Under the current economically stringent circumstances, it is practically untenable to catalog each document and therefore compromise would have to be made to catalog them on a collection level cataloging basis. This raises a new issue of whether it is worth spending that much effort when considering the effectiveness of limited access points. Each document contains specific information on specific topics and because of this, the document is valued. On the other hand, if many documents are grouped and given access by broad terms, it is difficult for users to have access to each document with its specific topic(s), unless the electronic database is maintained in some way on a permanent basis and provides cross references to the location within a group of documents.

In Japan, there is a center, an auxiliary to the Hosei University, which accepts various institutional collections of grey literature for control, maintenance and preservation. If there are several such centers organized in various parts of the world, this would be an ideal solution for the JDC and for other grey literature collections.

When the grey literature collection is part of a larger library, then how the collection is treated, given its ephemeral characteristics, must be determined by the institution. One issue discussed above touches on collection development, while another set of issues revolves around long-term treatment of the materials. Evaluation of cost versus benefit is an inevitable factor, particularly in context of the institution's planning and the realities of competing needs amongst its departments.

The second decision must be on the access, i.e., bibliographical control. The documents in printed format are to be processed according to the Library's normal practice in order to be kept in a more permanent and regular book stacks. This means that the documents must be cataloged and the bibliographic records must be added to the Library's online catalog. For that purpose, ideally each record must be created following the current cataloging rules, containing various access points according to the pre-established headings. Again, issues of human resources and costs must be seriously weighed against the benefits of having the materials and providing access to them.

Format of supplied information: printed versus electronic

From its inception, the Center's approach has been to collect printed documents. It is finding that an increasing number of documents in the grey literature category appear on web (in Japanese as well as in English). Starting in the mid 1990s, the Japanese government agencies have been making efforts that aim for wider distribution of their information, both for the Japanese public as well as the broader international public. Specifically, Japan's government documents on fairly specialized topics can be obtained in full text on the agencies' web. This is indeed a very welcome phenomenon. It would be extremely useful, if the following

information could be provided along with the reports and documents, even as a rough guideline:

- 1) Categories of the organization's information that are made available on web;
- 2) Timing of posting and how long the information will remain on the web, preferably in respect to the issuance of the printed format; and
- 3) Whether the information can be freely copied and disseminated for research purposes, or whether it is copyrighted

It would be helpful to the Center if we know which agencies are trying to put some information on web, and if possible, types of reports, whether full text or summary, presentation of graphs and charts, whether or not there are any restrictions, etc. Knowing this would eliminate redundant searches and the chase for printed versions on the part of the Center and its Tokyo Acquisitions Facility. At least the guidelines would give us a clue to decide whether or not to seek out the electronic version. If we know the timing of information being up on the web (in relation to the time of issuance of the printed document), that will give us good sense of locating information on the web in a timely way. This would also affect the decision of whether or not to seek printed materials, and yet enable us to be positioned to deliver the needed information to users within the time limits. Needless to say, it is also important to know how long specified information stays on the web.

This brings up a related issue. Most Japanese government agencies issue English translations of some documents. It should be noted that some are abridged versions or are short summaries. It helps to know the same information as above, such as which documents are translated, whether it is an abridged version, and how late these appear after the issuance of the original documents in Japanese. Those English translations are much sought after internationally and, therefore, Japanese government agencies must realize and respond to the needs accordingly. The author wishes to know other countries' practice.

Here is the Center's practice, which is as a related subject is certain to benefit users. When the Center provides a copy of a certain document, it sometimes finds an English summary prepared by the issuing agencies and listed on the web. We attach the print-out of those summaries to our original documents, and make a note in the bibliographic record of its availability as an attachment.

Japanese government documents are, with or without copyright notice, considered automatically copyrighted at the time of issuance. As for electronically disseminated information by the Japanese government agencies, an oral response to our inquiry was that it is open for everyone and can be freely copied. The Center's practice is that when it finds certain information on web that might be useful for research, it refer users to it but does not necessarily make a copy for the users.

Archiving

In relation to the above topic of electronically disseminated information, the author has been eagerly expressing the need for the Japanese government agencies to archive what they have put on web. While most individual organizations and companies do archive what they put on their web, there are fewer possibilities for the public to see their full texts; if the printed documents are reformatted to electronic form, the information is mainly for their organization's internal use and is generally not available for the public.

So far, the Center's finding is that the Japanese government agencies' full text documents on the web are in fact the electronic version of the printed documents, and it seems there has not yet been any documents that are issued in electronic format only. The author suspects that when such information is made available only in electronic form, it may be on Japan's special government local area network, such as the Kasumigaseki LAN and thus may not be available to the public, at least at the beginning. This is understandable because the information in those reports is, after all, issued mainly for the government's use to carry on the nation's business.

Again, the author's speculation is that under the current circumstances, provision of the electronic versions of various documents is an additional service to the public on the part of the agencies. The logic follows that since the printed version is the basis, the electronic version as a duplicate does not have to be archived. This assumption would have to change once the information is issued only in the electronic version, and it would be essential to provide archives and appropriate indexes. For the time being, whether to make a printed copy of electronically disseminated information for archival purposes for our own collection is something to be decided internally by the Center.

Conclusion

The above discussion covers a number of the current and important issues the Japan Documentation Center faces. Some are particular to the Center but others are of universal nature. In an ideal world, an information provider would be able to meet users needs by being able to obtain and provide that information in a timely way. The Center's users have proven to be diverse, with no major geographical concentration nor particular disciplines and having varying language abilities. Because of the collection being predominantly in Japanese, an approach was developed to maximize access to the collection through creating English language abstracts and indexing, and also by putting the database on the web. In order to reach users, and potential users, the Center continuously makes outreach efforts utilizing a variety of methods.

The Center collects the most recent public policy information in order to meet many researchers' need for timely information; the Center tries also to adapt its collecting focus to the political, economic and other changes in Japan in order to meet the changing research trends. Thus these characteristics of the grey literature collection bring up issues of how it would be treated, both bibliographically and physically, by the library institution of which this collection is a part, particularly on a long-term basis.

Finally, the Center has certain limitations stemming from practices of the issuing organizations. The challenge is to work with these outside factors so as to maximize those sources while minimizing redundant processing of the information.

There are yet other issues concerning grey literature, and the author wishes that they should be discussed vigorously in various forums so that better solutions could be found to benefit us all.

CONTROL AND ACCESS FOR GREY LITERATURE IN BRAZIL: A PROPOSAL.

ALMEIDA, Maria do Rosário Guimarães

ABSTRACT

Grey Literature in Brazil. Grey Literature is introduced in a constant transformation and incorporation of attributes, which makes it more and more different. There is a relation among its distinct classifications, definitions, characteristics and typology as of its importance. Grey Literature in Brazil is studied in accordance to its, visits, interviews and consultations of the database of the major Centres of Documentation and or Information in the Country. From this study, a proposal is made for the building of a Co-operative System of Grey Literature in Brazil, based on the System of Information on Grey Literature in Europe (SIGLE). System should be appropriate to our reality.

Key Words

Grey Literature; Technical Memory; Documental Memory; Institutional Memory; Technical Central File, Non-commercial literature; Non-conventional literature, Semi-published, Dark document, Invisible, Informal, Fugitive, Ephemeral.

1 INTRODUCTION

The transmission of information is limited to a reduced number of people, and it is also necessary to have printed methods of communication that may cover a large audition such as: journals, thesis, conferences records, written report, patents, databases, eletronic publications, CD-ROM, videodisc, including any message sent through the networks of communication

It is emphasized as the principal means of non convencional communication, the Grey Literature, accepted in almost all scientific areas. It is necessary to update the secondary services in order to obtain more efficient access to the documents produced. There is no doubt that Grey Literature covers a large and important amount of information, and for this reason it should be considered as the main source of information.

The term Grey Literature acquires relevant meaning and became common current use at the Seminary of York. That happened from 13 to 14 of December in York (United Kingdom). About this matter and, according to Bichteler, the expression was coined by the British librarians becoming of popular usage in 1980s.

Grey Literature includes important informations to a considerable number of customers. It is a fact that the volume of scientific material covered, today, by this kind of literature escapes the editorials circuits, and represents an extraordinary source of valuable inedit informations. Therefore its growth is based on economic reasons that sets forth, today, a starting point towards information.

2 THEORETICAL REFERENTIAL

Grey Literature, definitely, is not a new kind of document, whose form we cannot clearly precise, nor define all its contents. Thus it is a documentation with potential existence.

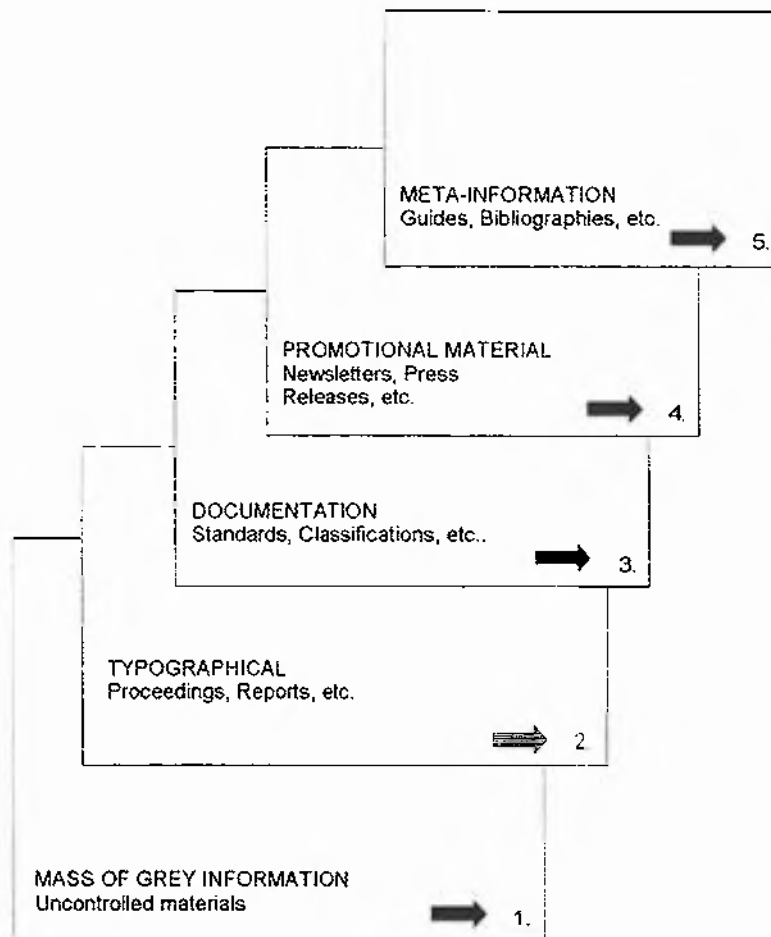
This kind of document is also known as non-conventional document, semi-published, dark document, invisible, informal, fugitive, ephemeral, and subterranean. It is characterized by a restrict circulation, and limited access and availability. This material is not submitted to any systematization process; it presents difficulties of bibliographic control and, consequently, difficult localization, reasons which make its acquisition economically expensive.

Grey Literature refers to documents of limited circulation which cannot be obtained through the conventional channels making its acquisition difficult. From this particularity, they are considered valuable documents and the most current examples are the technical and scientific report, thesis, private communications, recommendations, official documents, etc that need a commercial diffusion.

We should mention the experiments of Leonardo da Vinci, who used to elaborate written reports about his experiments, but never worried about publishing them. So, this way, we may consider them as grey, in the sense that they have not been published. This does not mean that Grey Literature is an indifferent object, without name, without its own characteristics. It appears as something latent, not real, at the time of the greatest illustrator of the Sistina Chapel.

Grey Literature has been in use since 1920 and according to Schmidaier, mentioned by Auger, it was called *Kleinschrifttum* - small literature. To our knowledge, the Germans had already had knowledge of this kind of literature through new publications, not available in libraries, but through the German National Bibliography since 1931.

Grey Literature is a subject that is in a constant transformation and incorporation of attributes, that makes it different from its own movements, by the inclusion of typical systems of meta-information or of abstracts. (see Figure 1)



Source: GREY Literature Network Service. NewsBriefNews, v. 5, n. 1, p. 2, 1996.
Figure.1 Evolution of the Grey Literature

According to WOOD¹ it is not convenient to define Grey Literature attending to this kind of material. Frequently, such documents are available to those interested in the time, lapse 12 and 18 months, including magazine articles as well as books.

The York's Seminary participants affirm that:

"...the essential step of this literature class, almost equivalent to so called non-conventional literature, is that it is not issued on the conventional circuits of commercial publications, and its access, is difficult in most cases. Furthermore, it's a literature that, in so many opportunities, has no intention of being published."²

¹ WOOD, N. David. Management of grey literature. In: DURANCE, C.J. (Comp.) **Management of record information**. Munich: Souer, 1990. p.61-62.

² Id. Ibid. p.62

BRAUN considers that Grey Literature:

"...is formed by scientific contribution, and also, by a considerable proportion of material whose character are informative and documental that, on frequently, has documental sources value. To BRAUN Grey Literature would be constituted by primary literature (written reports, thesis, and off prints), and also by other non-conventional materials such as: the international publications organisms, commercial chambers, banks, companies and official organisms"³.

To WOOD, it would be:

"...material which is not available through normal bookselling channels..."⁴.

ALVAREZ-OSSORIO, defines it as "... the whole of documents, that have varying typology, and are not published through the usual channels of scientific transmission"⁵.

Authors as CURRÁS and HENRICHS talk about documents:

"... that offers a primordial information, in an non-conventional shape and in most cases, without specific classification because it does not have an International Standard Serial Number (ISSN), International Standard Book Number (ISBN), Identification of Number of the Official Publications (NIPO), Legal Deposit (DL), and it cannot be acquired through the usual channels of diffusion"⁶.

The term Grey Literature was defined once more in the Third International Conference on Grey Literature which took place in Luxemburg, November/1997, as follows:

"that produced in a governmental, academical, business and industrial setting, in a print or eletronic form that is not controlled by the editorial channels"⁷.

³ BRAUN, H. Apud: MORENO-TORRES SANCHES, Rosario. El acceso a la literatura gris: actas de congreso y tesis doctorales. *Boletín de la ANABAD*, Madrid, v.36, n.4, p.673, oct./nov.1986.

⁴ WOOD, David N. Op. cit., p.61-62.

⁵ ALVAREZ-OSSORIO, José R. Pérez. *Introducción a la información y documentación científica*. Madrid: Alhambra, 1988. p.29.

⁶ CURRÁS, Emilia, HENRICHS, Norbert. *Estructuración de un programa de sistema experto e hipertexto para reconocimiento y localización de ciertos documentos de literatura gris*. Madrid, [1993]. Pretirada.

⁷ FARACE, Dominic. Grey literature and publishing. *NewsBriefNews*, Amsterdam: TransAtlantic/GreyNet, v.7, n.1, p.4, 1998.

The author defines Grey Literature, from all the definitions studied, as:

The set of documents notwithstanding its typology and support or electronic print form, emitted by universities centers, research centers, enterprises, industries, public and private academical associations, without the intention of being published and which are of vital importance for the transferring of knowledge.

It constitutes common aspects of the many definitions, access difficulty, circulation and the non-commercialization of such documents.

2.1 Grey Documents

When we report to the typology and the support of the grey documents, it becomes difficult to typify them in relation to the information, or the access and known area. We conceptualize some kind of grey documents. (see Table 1)

TABLE 1 Classification of kind of the Grey Documents

CLASIFICACION OF GREY DOCUMENTS	DEFINITIONS
PROCEEDINGS	Publications, or serie of publications, that contains the texts about expositions or communicated, orally facts transmitted in a conference, society or institution, proceeded of all the areas of knowledge.
NORMS	Documents that reflects in a consent of a community in a way to fix a model that should be practiced.
REPORTS	Documents that describes the development or the results of dominion of a tectonic or scientific research.
PATENTS	Documents that contain an invention which is guaranteed by the government by the government with a right also of product, use, sell, obtain and benefit itself for a determined number of years.
COMMERCIAL LITERATURE	Edited documents by manufactured or commercial enterprises with the finality of information or propaganda, that contain affable about products and services.
THESES (Ph.D.)	Original documents of research, presented to obtain the degree of doctor
BULLETIN	Documents have emanated from determined institutions or societies.
OFFICIAL PUBLICATIONS	Edited documents by a Government organism, with themes related by administrative,

2.2 The typology of Grey Literature

In ROSA DI CESARE's view point from the National Research Center (CNR) of the Central Library of Italy, Grey Literature is found distant from the conventional or white literature. The author uses the grey tones to classify such literature. (see Figure 2).

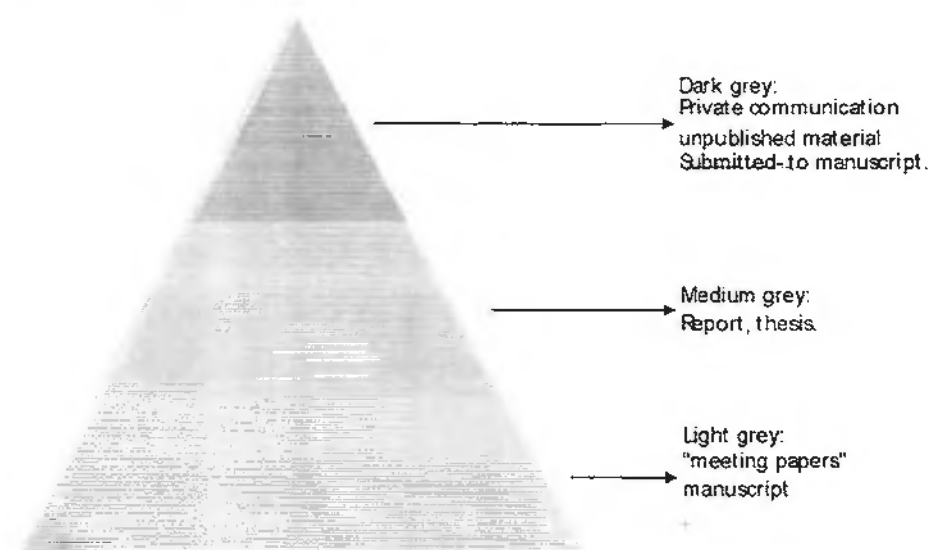


Fig. Classifications of Grey Literature by tons

2.3 Characteristics of Grey Literature

Grey literature presents specific characteristics which differentiates from the conventional ones. It has limited circulation, because it is constituted of few documents. It contains data to user which are quickly sent, and reach few readers.

The following peculiarities of this literature are pointed out:

- a) variable rules of editing and production, produced through difference output mechanisms of reproduction, including the most simple;

- b) ineficacious bibliographical control: the literature is not found in the banks of information agencies, that is, it is not submitted to ISSN, ISBN, LD and NIPO*. Therefore, it does not go through the commercial sources;
- c) difficulty in handling: much of this information is supported on microfilms and microfiles or other kind of material, where handling is difficult and turned down for consultation;
- d) conciseness: it has a less detailed character than the conventional literature, since it is not primarily for the purpose of publication. As mentioned before, in general some months later it is published by the commercial channels, in the form of magazine articles or books, because of the need to make the information's known is a necessity;
- e) institutional production: as a general rule, Grey Literature is the responsibility of government institutions, universities, and research centers. Finally, it is almost always published by an institution.

What we can see, in an international scale, is a tendency in which the selection, the analysis and the treatment of the grey documents follow the standardized systematization of in/out registration, in the means of having a precise, national and international registration aiming at the availability and accessibility of the material without any problem on the course of going from the producer to the user. These works in a net system are made in collaboration with the researcher, the library science or information experts.

Therefore, for the treatment of the grey material, some agencies that follow the rules based on the norms of International Standard Organization (ISO), the Anglo American Cataloguing Rules (AACR2), the Classification of Dewey, of specific thesauri in the areas of performance, descriptions of documents according to the rules of the Committee on Scientific and Technical Information (COSATI).

Another aspect to have in mind is the technological transference. There are many ways and requisites to develop. However, the more effective transferences

* This characteristic has gone through changes since the 80's

occurs when fundamented in the interchange of personnel or in the interpersonnel transmission through the new world of micro computers and the information network, that offers new potentiality, as are the electronics publications, visualizing technics, virtual reality, among others, with the growth of the information industry.

2.4 The electronic grey document

Experts on this theme, worried in giving continuity to the cooperation among the organizations aiming at providing a better access to the world of interconnected information, that it is necessary to know well the systems of hypertext to obtain information from other sources.

The interconnected information is not a linear extension of the printed information. The fast developement of an informative and interconnected world, and, specially, with the Internet has been having a giant impact on the access of information and how we deal with this new faceted, bringing great advantages to the Grey Literature, as well as one of its charateristics which is the absence of bibliographical control.

As to the volume and usage of the grey publication, these play and important role on the scientific information. Examples are mentioned at the International Conferences of Amsterdam in 1993 (GL'93), Washington, D.C. (G L'95) and Luxemburg (GL'97). The next Conference will be held on October 4 and 5, 1999 in the city of Washington, D.C., and will have the following theme: The new frontiers of the Grey Literature.

It is concluded that, with the development of Internet, new kinds of electronic grey documents are ocuring, such as: the electronic printed mail, the grey files and all the technical documentation retrived through WEB. This new kind of Grey Literature is characterized by the fact that it is a combination of documents that are not textually complete making it less understood. Through new hyperlinks, it is permitted that new kinds of Grey Literature become virtuality unlimited.

It is verified that the grey information constitutes an important part on the scientific and technical communication. The grey documents remain unique and are rarely followed by a printed article. Part of the Grey Literature is published by well known institutions whose names guarantee its quality.

With regards to the production increasingly voluminous of Grey Literature, it is estimated to be 3 to 4 times greater than the conventional or commercial one, according to DI CESARE⁸

"...Internet offers new means of scientific and technical communication, and new paradigms of communication are emerging.[...] the world information changing and electronic information sources, and particularly, Grey Literature are increasing. In this way, problems related to bibliographic control, will not be solved only by using new technologies. Therefore appropriate standards for successful identification and easy access to desired information will be more necessary."⁹

2.5 The production of Grey Literature in Brazil

Through bibliographical search in database and Brazilian Libraries, we detected a strong tendency of typological literature that registers the production of Grey Literature such as: Technical Memory, Institutional Memory, Documental Memory or Technical Central File.

Along with this case, the necessity of establishing a terminology for the Grey Literature and the need for norms to standardize the treatment for this literature was detected.

A nationwide study of the main institutions was carried out. A list was produced per State and per Institutions within each State (see Table 2)

⁸ GREY LITERATURE NETWORK SERVICE. NewsBriefNews. Amsterdam: TransAtlantic, v.6, n.1, p.2, 1997.

⁹ DI CESARE, Rosa, SALA, Cesare. The use of grey literature in the agricultural economics field: a quantitative analysis. In: INTERNATIONAL CONFERENCE ON GREY LITERATURE, 2., 1995, Washington. *Conference Proceedings...* Washington, D.C.: TransAtlantic/ GreyNet, 1995. p.168.

TABLE 2 Institutional sources/Brazil

INSTITUTIONS	SIGLE	CITY
National School of Public Administration Foundation	ENAP	Brasília
Agricultural Documentary Informational Coordination	CENAGRI	Brasília
National Center of Environmental Information	CNIA	Brasília
Science and Technology Information Brazilian Institute	IBICT	Brasília
Documentation and Information of the Federal Senate Undersecretary	SDISF	Brasília
Latin America and Caribbean Health Science Information Center	BIREME	São Paulo
Technology Research Institution of São Paulo State	IPT	São Paulo
Technology Information Sector	CETEC	Belo Horizonte
Nuclear Information Center	CIN/CNEN	Rio de Janeiro
Information and Development Center of Petrobrás	CENPES	Rio de Janeiro

A research was performed within huge national information centers and/or documentation, pointing out the following areas of knowledge: Health, Public Administration, Law, Politics, Environment, Nuclear Energy, Agronomy, Politics and Sciences of Information, among others; the production and usage of Grey Literature in Brazil was quantified this way.

In the case of Brazil, isolated management for Grey Literature were detected, but it was proved the lack of a co-ordinating organism that would unite and make known this kind of literature, in order to speed up the accessibility and availability of the grey documents.

It is worthy of comment that the only Brazilian experience on Grey Literature, is from the co-ordination of Professor Dinah Aguiar Población:

"...since 1992, the National Association of Research and Post-graduation in Information and Library Science - ANCIB) concentrated its investigation activities in integrated projects that are composed by many subprojects supported by the National Council of Research and Development (CNPq). Two of these subprojects develop research in scientific production, considering Grey Literature as all kinds of documents created with non-conventional characteristics. The thesis and events that constitutes the object of the subprojects were selected. That research had the objective identify the production and the producers characteristics of the communications that were

presented on the events taking place in Brazil..."¹⁰

In this sense, the problem to be studied will have an empirical object to establish a system to administrate the Grey Literature in Brazil, considering the importance of this material for the scientific and technological development. In spite of the co-operative systems, it was observed in Brazil the need for a system turned to the treatment and dissemination of Grey Literature.

Keeping this need in mind, the study intends to incentivate the Brazilian institutes to try to solve this matter, highlighting the manegment of policy based on the System for Information on Grey Literature in Europe (SIGLE).

For the establishment of such system, a proposal has been formulated and it will be detailed as follows.

3. PROPOSAL FOR CONTROL, ANALYSIS AND TREATAMENT, DISSEMINATION AND USE OF GREY LITERATURE IN BRAZIL

Grey Literature is the object of concern in the Centers of Documentation and Information, Libraries, Research Institutes and others, due to the difficulties of access found at the time of colleting the material in the institutions which study, analyse and produce this literature.

In view of the international experience of cooperation among information centers which work with the management and dissemination of Grey Literature, it is proposed that this system should follow the structure of the System for Information on Grey Literature in Europe (SIGLE); adjusting it to our reality.

¹⁰ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO E BIBLIOTECONOMIA. *Eventos em arquivologia, Biblioteconomia, Ciência da Informação, Documentação e Muscologia realizados no Brasil: 1951-1994*. São Paulo, 1994. p.1-2.

3.1 Objectives

3.1.1 General objective

Building up the structure of a co-operative system to manegment Grey Literature in Brazil is such a way as to make it possible to identify its writers; the establishment of rules and technical standards to control, the dissemination or spreading and transferring of grey information in a national and international setting.

3.1.2 Specific objectives

a) Identify the units which produce Grey Literature in Brazil.

In order to obtain such objectives, registering units wil have to be built for those who produce Grey Literature and integrate the Brazilian co-operative system, to make it possible to adminstrate. The system should incorporate some units of work located in different kinds of institutions such as:

Academies	Hospitals
Foundations	Institutions
Laboratories	Museums
Observatories	Educational Centres
Societies	Investigation Institutes
Governement and no- government	Universities
Agencies	

b) Establish rules and standards of bibliographical description and compensation for Grey Literature, compatible with other existing systems.

The rules and standards of bibliographical description and the compensation of Grey Literature can follow the guideline and manuals

adapted by SIGLE, internationally, well known through the descriptive cataloguing rules based on the International System of Nuclear Information (INIS).

To classify the themes, a modified version is adapted from the original classification of the Scientific and Technical Information Committee (COSATI) which belong to the United States Science and Technology Federal Council. And it is also known by the name of COSATI classification system (classification composed of letters and numbers). This classification is in use by SIGLE.

- c) Establish rules for recovering, analysing, dissemination and use Grey Literature produced in Brazil.

The rules will be established to the integrated and non-integrated units of the system, detailing its recovering, treatment and dissemination of the literature in order to systematize the services related to the grey documents, so it can be sent to the coordinator unit because it is a decentralized co-operative system that acts as such decentralizing the tasks and coordinating them.

- d) Create mechanisms to make possible the interface communication with similar systems of production, treatment and dissemination of Grey Literature.

The mechanisms which allows the communication's interface consist of having the electronic support and the knowledge on the system to be implanted. The recovery of the information to optimise the physical or the electronic access of the grey documents is expected from this procedure.

4 THE METHODOLOGY

4.1 Scope of the research

It is constituted of the material available in : Academies, Foundations, Laboratories, Observatories, Societies, Government and non Government Agencies, Hospitals, Institutions, Museums, Educational Centers, Institutes of investigation and Universities in Brazil

4.2 Procedure

- a) contact with the several institutions which produce Grey Literature;
- b) localization and retrieval of the available literature;
- c) doing the collection of Grey literature;
- d) identify the kind of material and support of Grey Literature;
- e) develop a program of database elaboration;
- f) find out the scientific production of researchers and
- g) quantify, through bibliometric studies, the Grey Literature produced in Brazil.

5 SYSTEM MANAGMENT STRUCTURE

The manegment of the system can be reached by an Institute that has network experience. There is the Brazilian Institute of Scientific and Technical Information (IBICT) as an example. It is composed by other units and forms a new administrative body of the co-operative system. The folowing it is recommended:

® a National Coordinating Center, as in the IBICT case. In all co-operative centers, it is necessary to designate an organ or organs to take care of the different activities.

® a technical group is composed of representative(s) of Co-operative Centers from the National Network. This group should be in charge of the technical part of the national system, so it can gives a better efficacy of their services and products.

® an advising committee integrated by representatives from Education and Culture Ministry (MEC), National Association of Research and Post-Graduation in Information and Library Science (ANCIB), Getúlio Vargas Foundation (FGV). These are prestigious and capable institutes to guide the system in its totality.

® Co-operative Centers will be comprise all the units that produce and compose part of the National Coordinating Center.

® a GreyNet representative, being the most important network in dealing with grey material.

For the development of Grey Literature, it is necessary the promotion and dissemination of event calendars, seminars, workshops, the incentive of electronic report editions, press reports, the creation of web pages and networks in the institutions which are working with Grey Literature, the production of secondary or meta-information sources.

Due to enourmous economic crisis that is reflected upon the scientific production, the proliferation and the use of Grey Literature can be evidenced. It is constituted as a positive factor to the administration of such proposal in Brazil. That would greatly benefit students, as well as teachers, investigators and other professional fields which are growing with the advent of Internet.

Even though Grey Literature is an important source of information that can reach scientists and researchers in a quick and efficient way, the transmission of information is limited to a reduced number of people. Printing methods for the transmission of information, that can reach a great audience, such as: journals, thesis, conference records, reports, patents, rules, review on specific topics, summaries, indexes and database.

It seems necessary to discuss about the importance and effectiveness of this system. The materialization of this initiative, no doubt, will be

an action leading to what includes its recognition, speeding up its production and, consequently, the establishment of Grey Literature.

By means, of what was stated above, we can see how opportune it is a system of such importance to Brazil, specially because of the few access to precise and concise information. The current proposal shall be implemented aiming at covering faults in the treatment and access to Grey Literature, giving concrete and significant elements to fortify the activities inherent to the researchers, librarian and library science and information professionals in Brazil.

BIBLIOGRAPHY

- ALMEIDA, Maria do Rosário Guimarães. La Literatura Gris: sistemas y redes en el ámbito nacional e internacional. Una propuesta para Brasil. Madrid: Universidad Complutense de Madrid, 1998. 573p. Tese (Doutora em Ciencia da Informação) – Departamento de Biblioteconomia e Documentação. Universidad Complutense de Madrid, 1999. 547p.
- ÁLVAREZ-OSSORIO, José R. Pérez. *Introducción a la información y documentación científica*. Madrid: Alhambra, 1988. 107p.
- ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO E BIBLIOTECONOMIA. *Eventos em arquivologia, Biblioteconomia, Ciência da Informação, Documentação e Museologia realizados no Brasil: 1951-1994*. São Paulo, 1994. 128p.
- ASSOCIATION FRANÇAISE DE NORMALIZATION. *Vocabulaire de la Documentation*. Paris, 1983. [Projet de norme enregistrée].
- BICHTLER, Julie. Geologist and gray literature access, use and problems. *Science & Technology Libraries*, v.11, p.39-50.
- CARROL, Bonnie C., COTTER, Gladys A. A new generation of grey literature: the impact of advanced information technologies. In: INTERNATIONAL CONFERENCE ON GREY LITERATURE, 1., 1993, Amsterdam, *Conference Proceedings...* Amsterdam: TransAtlantic, 1994. 333p.
- COMBEROUSSE, Martine. Les nouvelles technologies au service de la littérature grise. *Bulletin des bibliothèques de France*, Paris, v.40, n.2, p.51, 1995.
- CURRÁS, Emilia, HENRICH, Norbert. *Estructuración de un programa de sistema experto e hipertexto para reconocimiento y localización de ciertos documentos de literatura gris*. Madrid, [1993]. Preprint.
- DI CESARE, Rosa, SALA, Cesare. The use of grey literature in the agricultural economics field: a quantitative analysis. In: INTERNATIONAL CONFERENCE ON GREY LITERATURE, 2., Washington, 1995. *Conference Proceedings...* Washington, D.C. Transatlantic/ GreyNet, 1995. 333p.
- DI CESARE, Rosa. The use of literature in the agricultural economics field: a quantitative analysis. In: INTERNATIONAL CONFERENCE ON GREY LITERATURE, 2., 1995. Washington, *Conference Proceedings...* Washington: TransAtlantic, 1995. 333p.
- FARACE, Dominic. Grey Literature and publishing. *NewsBriefNews*, Amsterdam, TransAtlantic/ GreyNet, v.7, n.1, p.4, 1998.
- SCHMIDMAIER, J. Ask no questions and you'll be told no lies: or how we can remove people's fear of grey literature. *Libri*, v. 36, n.2, p. 98-112, 1986. Apud: AUGER, Charles Peter *Information sources in grey literature*. London: Bowker, 1994. 170 p., p.6
- WOOD, David N., SMITH, A W. SIGLE: a model for international co-operation. *Interlending & Document Supply*, Boston Spa, v. 21, n.1, p.22, 1993.

INIST : Tracking grey literature in a changing environment

by Joachim Schopfel

Institut de l'Information Scientifique et Technique du Centre
National de Recherche Scientifique INIST/CNRS :

Ladies and gentlemen, Mr. Chairman,

I'm glad to be with you today and to represent INIST, the Institute for Scientific and Technical Information in Nancy, France. There are many French projects concerning non-conventional literature but to this day, there has been no clearly defined policies. In other words, we have to deal with " low policy, low economics ". INIST is an element in a greater national and international context. Our contribution has to be set within the rapidly changing context of documentation and information. But, even if our situation may be slightly different from other institutions, there are at least some common problems and issues.

(1) INIST

First, let me give you some basic information on our institute :

INIST, the French institute for scientific and technical information, is one of the major European scientific information centers. Our goals are to collect, process and disseminate the results of scientific research. Our current acquisition policy is to acquire all material likely to be used by the research, higher education and industry communities, regardless of subject and language.

Nevertheless, as a part of the CNRS, the French national center for scientific research, our most important mission is to work for and with the French scientific community i.e. the main public research organizations, universities, industrial R & D units and so on.

Our principal activities are database production (Pascal and Francis), document supply and scientometrics research. Our databases contain nearly 15 million bibliographic records in science, technology, medicine, social sciences and humanities. Our library covers about 24 km of shelf space. The major part of INIST document resources are serial publications. About 9,000 serial titles are currently received by INIST, with another 14,000 titles in our holdings. 400,000 articles are digitized every year.

Copies from all these documents can be ordered from our online service platform (www.inist.fr). The documents which INIST does not possess are supplied via a network of more than 100 European referral libraries.

(2) Grey literature at INIST

Our “non-conventional” collections include about 100,000 doctoral dissertations, 60,000 conference proceedings and papers, 56,000 scientific reports. Currently, we receive more than 10,000 grey documents each year.

In the changing context of database production, e-journals and the evolution of the web, INIST is engaged in a collective reflection on its acquisition policy, services and products. I would like to outline some of the perspectives for non-conventional literature.

1. **Electronic documents:** Until now, we have limited our acquisitions to print documents, spending much time on traditional library activities like cataloging. Today, most non-conventional literature exists in electronic format. Therefore, we are re-orienting our grey acquisition policy to an “all-electronic” policy. Examples : some weeks ago, INIST started to encourage its partners to submit report or dissertation files by email ; we are working on a web portal in order to simplify submission of e-documents ; there are national projects for electronic dissertations and theses in which INIST could play a major role. And of course, we are engaged in “metadata discussions”.
2. **Access providing:** Until now, we have owned in our collections most of the documents needed for our databases and document supply. Today, this approach has reached its limits, as the example of NTIS shows. There are too many documents, on the web and elsewhere, for us to deal with, considering our limited financial and technical resources. In the era of the Internet, we do not need to have all the documents in our holdings. Instead, we have to know where to get them. This means

a systematic search on the web and a direct relationship with the main producers of grey literature. Here, I hope the GL compendium project will be helpful.

3. **Consulting and assistance :** Down the road, this also means assisting grey literature producers, in order to facilitate the search for this special type of documents. Two years ago, INIST became the French national ISRN agency. The International Standard Report Number contributes to the identification of report producers ; it facilitates document management and identification in databases and catalogs; it may also contribute to the creation of URNs. Last but not least, we are contributing to the standardization of report presentation.
4. **National public research :** I said that we did not need to have all the documents in our holdings. That is true. But there is one exception. INIST plays a specific role for French public research. We produce the science and technology part of the national thesis database, we collect scientific reports from public administrations and other research organizations, and proceedings of the major French conferences. La Documentation Française proposes an access to the administrative reports via an online database; currently, we are exploring the opportunity of a public STM report database on the web, together with other partners.
5. **CNRS database:** Today, we are also involved in a project of a CNRS database of the publications from the CNRS research units. Tomorrow, this may include pre-prints and other unpublished papers. Creating and providing a sort of web gate to these documents is a part of our role as keeper of our “documentary heritage”. In the future, such a database will also be a kind of guarantee against “dirty-grey documents”.
6. **Links:** It may be possible to link this documentary database to information on French experts and research units in science and technology.
7. **European level:** Facilitating access to French research production includes participation in the European EAGLE network. INIST is the French input center for the SIGLE database. Next time, I think we should discuss the future of this database in the context of the internet, the full text, and free access.
8. **Education and training:** The changing landscape of information and documentation calls for new skills and abilities. INIST is implementing an expensive program for the training and development of its workforce, together with ENSSIB, the major French

Library Science school at Lyon, and other institutions. INIST participates in the higher education of documentation specialists; each year we welcome trainees for research and other projects, we acquire French dissertations in information science, and we are creating at INIST a kind of “ library-laboratory ” linking higher education, user training, documentation, informatics and research. Assuming its role in the national research context, INIST will organize in 2000 the 2nd conference for the CNRS documentation specialists where grey literature will certainly play an important role.

To conclude, I would like to insist on the dual function of INIST in knowledge mediation :

- on the one hand, we provide and facilitate access to grey literature wherever it may be, on the web or elsewhere ;
- on the other hand, by archiving French grey literature, we act as keeper of a special document heritage,.

We all know that the lifetime of non-conventional literature is rather short. In the era of electronic documents and in a culture of ephemerality, it may be necessary to save at least some of these documents from oblivion and extinction.

The Competitive Advantage for the New Millennium:

Knowledge Management

Atefeh S. McCampbell, Ph.D.

Linda Moorhead Clare

Scott Howard Gitters

Florida Institute of Technology
Building 5442
Aberdeen Proving Ground
Aberdeen, Maryland 21005

ABSTRACT

The following paper defines the newly emerging concept of knowledge management. The topics presented include: principles and practices of knowledge management and the competitive advantage gained by industries employing this discipline. The technical applications and tools currently utilized within this discipline will also be discussed.

A survey was conducted to quantify the level of corporate awareness of knowledge management, the effectiveness of knowledge management practices on leveraging intellectual capital, the measurement of knowledge management initiatives and the technology employed to achieve those initiatives. The survey was distributed to the following industries: manufacturing, transportation, communication, financial services, retail trade, healthcare, government, and business and professional service organizations. The results of the survey indicated a majority of companies are employing knowledge management practices with positive effects on profitability, as well as, customer satisfaction and quality of service. According to the survey, technology solutions, inclusive of, smart documents and networks, are playing a pivotal role in the successful implementation of knowledge management strategies.

Knowledge Management Case Studies are also included on the following firms: Teltech, Ernst & Young, Microsoft, and Hewlett Packard. These case studies are analyzed to determine the competitive impact of knowledge management strategies.

Finally, conclusions are drawn regarding the strategic direction of this new discipline and its effect on competition, productivity and quality for the business of tomorrow.

INTRODUCTION

Knowledge Management-the Competitive Advantage

The study of knowledge management evolved from the need for companies to manage resources more effectively in a hyper-competitive, global economy. Ikujiro Nonaka stated in The Knowledge Creating Company, "In an economy where the only certainty is uncertainty, the one sure source of lasting competitive advantage is knowledge... Successful companies are those that consistently create new knowledge, disseminate it widely throughout the organization, and quickly embody it in new technologies and products" (Nonaka, 1991).

The American Productivity & Quality Center (APQC), a nonprofit education and research organization which fostered the creation of the Malcolm Baldrige National Quality Award, defines knowledge management as "the strategies and processes of identifying, capturing and leveraging knowledge" to enhance competitiveness (Manasco, 1996).

As the intellectual assets of companies become more complex, the ability to inspire employees to invest in knowledge will become more, not less, important. (Leonard-Barton et al., 1995).

According to Gartner Group findings for 1998, implementation of knowledge management systems has begun, and is currently being deployed by most large companies. One third of Fortune 1000 companies are now including knowledge management initiatives in their 1999 plans (Smalley-Bowen, et al. 1999).

According to a study of successful knowledge management projects, knowledge management was at least partially responsible for a major transformation of one large consulting firm. The transformation was extensive in terms of a marked improvement in financial results for the firm while engaged in knowledge management practices. Line consultants drew heavily from the firm's centralized knowledge centers, accessing previous presentations to other clients, process and system design specifications, work plans and other project-oriented collateral and artifacts. Senior managers described knowledge management as the core of the consulting strategy and the concept was pervasive in the company's internal and external documents (Davenport, et al., 1998).

According to Yogesh Malhotra, author of Deciphering the Knowledge Management Hype (1998), "Knowledge Management caters to the critical issues of organizational adaptation, survival and competence in face of increasingly discontinuous environmental

change... Essentially, it embodies organizational processes that seek synergistic combination of data and information processing capacity of information technologies, and the creative and innovative capacity of human beings" (Malhotra, 1998).

The new world of knowledge-based industries is distinguished by its emphasis on precognition and adaptation, in contrast to the traditional emphasis on optimization based on prediction. Arthur suggests that the new world of knowledge-based business is characterized by a continuous redefinition of organizational goals, purposes, and an organization's "way of doing things". This new business environment is characterized by radical and discontinuous change and demands anticipatory responses from organization members who need to carry out the mandate of a faster cycle of knowledge creation and action based on this new knowledge (Arthur, 1994).

With companies increasingly being viewed as mechanisms for coordinating knowledge, the mechanisms that companies utilize to integrate the specialist knowledge of individual members into goods and services become increasingly important (Grant, 1996).

A recently published 1999 Executive Briefing, *eBusiness Imperative*, by The Concoors Group concludes that the ability to deploy knowledge is the principal differentiator in the Information Age. The emerging 21st century information age enterprise is anything but linear, rather growing outwardly. Businesses are competing on the basis of business models with brief half-lives. Products and services are differentiated by knowledge and intellectual capital. It is clear, in the move from the industrial age to the technological age, intellectual property or knowledge impacts profitability.

Principles of Knowledge Management

Knowledge Management is an emerging discipline which is a dynamic process. According to a study conducted by American Productivity & Quality Center's International Bench-marking Clearinghouse, of eleven organizations participating in the arena of knowledge management: *Arthur Andersen; Chevron Corporation; Dow Chemical Company; Hughes Space & Communications; Kaiser Permanente; Price Waterhouse; Sequent Computer Systems; Skandia AFS; Texas Instruments; USAA; and National Security Agency*, the following was recognized:

1. The practice is dynamic, but often starts by creating, finding and collecting internal knowledge and best practices.
2. Sharing and understanding those practices so they can be used.

3. Adapting and applying those practices to new situations.

Also noted in this report was the direct reflection of top down management enabling, encouraging or undermining the process. Strong professional ethic and pride, supported by skills in cross- functional teaming, cultural support such as reward and recognition programs and CEO's as visible advocates for the sharing of knowledge and best practices, resulted in the promotion of success.

The most widespread strategy among the participants in the study was the transfer of knowledge and best practices in order to improve operations or to embed them in products and services. This includes systems and practices to obtain, organize, restructure, warehouse or memorize and distribute knowledge. Most of the firms also underscored the importance of teams, relationships, and networks.

These three elements are the basis for the effective transfer of knowledge. The study showed that the participating firms had taken part in a number of approaches to encourage collaborative knowledge transfer (Manasco, 1996).

Highly satisfied employees drives customer capital creation. Research has proven that skilled and motivated employees correlates with increased customer satisfaction and loyalty which further correlates to increases in market share and profits (Stewart, 1997). The team, relationship and network approach encourages employee satisfaction and motivation.

Conventional job descriptions mask the ways in which people actually work, but they also ignore the learning and innovation that are generated within informal communities of practice. The key to harnessing innovative energy lies in the effective design of organizations and the ways communities are linked (Seely-Brown et al, 1991).

The more common type of success in knowledge management involves operational improvements limited to a particular process or function. According to a study reported in the Sloan Management Review, the following factors lead to knowledge management project success:

- Link to economic performance or industry value
- Technical and organizational infrastructure
- Standard, flexible knowledge structure
- Knowledge-friendly culture
- Clear purpose and language
- Change in motivational practices
- Multiple channels for knowledge transfer
- Senior management support

Knowledge management is one of many components of good management. Sound planning, savvy marketing, high-quality products and services, attention to customers, efficient structuring of work and thoughtful management of an organization's resources are all critical to compete in today's marketplace. Knowledge management may help create the competitive edge in today's global environment (Davenport, et al, 1998).

Several recent studies report results that solidly show knowledge management initiatives growing in large corporations, up more that 100 percent from 1998, encompassing more than half of all major enterprises (Petch, 1999).

Methodology

A total of one hundred questionnaires consisting of seven questions were distributed to the following industries: manufacturing, transportation, communication, financial services, retail trade, healthcare, government and business and professional service organizations. The purpose of the survey was to quantify the level of corporate awareness of knowledge management, the effectiveness of knowledge management practices on leveraging intellectual capital, the measurement of knowledge management initiatives and the technology employed to achieve those initiatives. The survey also included demographic questions regarding industry type and respondents position within the organization. Analysis is conducted to determine existing and projected trends.

Summary of Findings

According to the survey, if knowledge management is defined as the strategies and processes to capture, leverage and disseminate tangible and intangible knowledge, 74% of the companies surveyed participate in the practice of knowledge management. The data indicated that 26% of the respondents did not participate in the practice of knowledge management.

According to the 74% of respondents who participated in the practice of knowledge management initiatives, 100% responded that knowledge management increased quality and productivity within their respective organizations. This is a clear indication of the highly successful outcome of knowledge management strategies across industries. Also, the improvement of quality and productivity within these organizations, fosters the competitive edge in the global business arena of today.

Effectively leveraging intellectual capital proposes a unique challenge. The survey indicated that knowledge management practices had varying effects on this issue. The data reports that 10% of respondents indicated that knowledge management practices had *no effect*, 28% of respondents indicated *some effect*, 39% of respondents indicated that their organizations knowledge management practices were *effective* at leveraging intellectual capital, while 18% of respondents indicated *very effective* and 5% indicated *extremely effective*. An analysis of these statistics results in the finding that over 60% of those surveyed, believe that their organizations knowledge management practices are, at the very least, effective at leveraging intellectual capital. This indicates a positive trend within the discipline to capture and leverage tacit knowledge. As knowledge management practices improve within an organization, one might predict an increase in this trend.

The survey raised the question of how organizations measure the success of knowledge management initiatives. The data reports that 13% of the respondents utilize ROI as the measurement tool. 36% of those surveyed, indicated Product/Service Innovation was the measurement indicator. Matching that percentage rate of 36% was Improved Profitability, while 29% indicated a variety of measurement tools inclusive of:

- ♦ Quality of Service
- ♦ Production Goals
- ♦ Customer Satisfaction
- ♦ Minimized Waste and Work Redundancy
- ♦ Employee Turnover
- ♦ Performance Improvement
- ♦ Cost per Access Line
- ♦ Process Cost
- ♦ Market share increase

3% of the respondents indicated no form of measurement was utilized. These findings indicate the trend to utilize not only measurement tools which reflect on the bottom line, but also, innovative concepts to measure the value of knowledge management implementation. In the opinion of the authors, further refinement of measurement techniques will result in a larger number of organizations participating in knowledge management practices.

The use of technology solutions to employ knowledge management practices is widespread across all industries. The survey indicated the following results regarding technology employed: Knowledge Process Mapping 15%, Smart Documents 15%, and

Networks 68%. 18% of respondents indicated the use of other forms of technology to implement knowledge management inclusive of:

- ♦ Technology-based Training Sessions
- ♦ Intra/Internet Knowledge Base for exchange of solutions by both client and employee

The use of networks, according to the survey results, is an integral facet of the successful capture and exchange of knowledge, not only internally, but externally. Respondents indicated the use of networks for research, strategy documents, training, flowcharts and organization-wide Smart document applications. As security is enhanced on the internet and firewalls improve, one might predict increased usage of internet based knowledge management data repositories world-wide.

Following are the demographic results of the survey by industry category and position within the organization:

Industry: Manufacturing 5%; Transportation 2%; Communication 2%; Financial Services 5%; Retail Trade 7%; Healthcare 39%; Business and Professional Services 23%; and Other (inclusive of Government, Construction, and Publishing) 13%.

Position: Chief Executive Officer 5%; Chief Information Officer 2%; Management 65% and Staff 26%.

Knowledge Management and Information Technology

As shown in the results of the survey, a key enabler for the implementation of knowledge management is information technology. IT's role is emerging as an integrator of communications technology, rather than solely a keeper of information. The critical role for IT lies in its ability to support communication, collaboration, and those searching for knowledge and information, not static repositories of "best practices", according to the American Productivity and Quality Center International Benchmarking Clearinghouse study (Manasco, 1996).

Karl Erik Sveiby, the author of The New Organizational Wealth: Managing and Measuring Knowledge-Based Assets, contends that the confusion between knowledge and information has caused managers to sink billions of dollars in information technology ventures that have yielded marginal results. Sveiby asserts that business managers need to realize that unlike information, knowledge is embedded in people, and knowledge creation occurs in the process of social interaction (Sveiby, 1997).

Knowledge can be managed as a strategic resource by focusing on the value of information technologies in exploiting organizational knowledge (Earl, 1994).

New design precepts are emerging as we accumulate experience. Participants interacting with the computer need to be able to interpret what's happening on the screen in terms of the real world. A well-designed learning computer lab leaves its participants with the skill to communicate without depending on a computer (Sense, 1994).

Innovative information technology is applied to support and facilitate organizational learning processes involved in the development and exchange of organizational members underlying opinions, assumptions and interpretations of the environment (Hine, M. et al, 1998).

There are four constructs linked to organizational learning: knowledge acquisition, information distribution, information interpretation and organizational memory (Huber, 1991).

The current business arena requires the need for a vast and complex interpretation of information outputs generated by computer systems. This variety is necessary to encompass the multiple global views of an unpredictable future. Brook Manville, Director of Knowledge Management at McKinsey & Company in Boston, views the implementation of these issues in terms of the shift from the traditional emphasis on transaction processing, integrated logistics, and workflows to systems that support competencies for communication building, people networks, and on the job learning (Malhotra, 1998).

A lot of intellectual capital resides in the minds of IT workers. Companies such as Andersen Consulting, Ford, and Monsanto encourage employees to put "tacit" knowledge, the know-how in their heads, into "explicit" form, such as written reports or video presentations. This captured knowledge is then stored in repositories such as databases and intranet Web servers, all of which users can search. The ownership of intellectual capital has become a priority for many organizations, with several recent court cases having thrust this issue to the forefront. Wal-Mart Stores, Inc. filed a lawsuit last year against Amazon.com that alleges theft of trade secrets. Wal-Mart claims the defendants carefully recruited members of its staff who, as a group, have the knowledge to replicate information systems and business processes Wal-Mart has spent years and financial resources developing. A recent Supreme Court action affirmed that business methods linked with software can be patented. Savvy technology managers now use a combination of legal defenses and common sense to protect their companies' most innovative IT systems and to retain their talented information technology staff (Hibbard, 1999).

Increasing the number of new patents, accelerating to retirement obsolescence of patentable ideas, decreasing the time to market for new goods, eliminating waste by consistently applying best practices and investment in information systems, training and education are all practices which add value and measurement of knowledge management (Edvinsson, et al, 1997).

Despite initial fragmentation, knowledge management technologies are quickly evolving and converging, spurred by requirements of top global organizations, attention by consultants and integrators and efforts by pioneering vendors (Mantelman, 1999).

Most knowledge applications have evolved from pre-existing types for managing documents, databases, workgroups and customers. Most explicit knowledge lives in documents, Web or PC files, paper or scanned images. Smart systems do more than track or store information. They help organizations manage content-in the context of what people know and need to know. These documents are now known as "Smart Documents".

Knowledge management requires collaborative group support. Software tools are available which let people build communities and take part in virtual teams; brainstorm, develop, present and deliver knowledge; share documents or applications; discuss and manage projects; and coordinate activities.

Teltech: The Business of Knowledge Management Case Study

Teltech was formed in 1984 and offers instructive lessons to companies wishing to better manage their knowledge and information assets. The company has built a successful business on helping companies gain access to external technical expertise and information. Teltech's vision includes the view that purely technical approaches to information and knowledge provision will rarely add value such as the incorporation of the human element. The knowledge analyst assists in defining what information is desired, clarifies concepts and terms, interprets search results and knows when and where to seek further information.

A key premise of Teltech's business model is that people are not only guides to information, but important repositories of expertise. A key focus of the firm's product line development has been the development of software providing an integrated view of sources of information on a particular topic known as an integrated source map. It will present virtually all information that a customer may want on a particular topic in a matrixed environment (Davenport, 1997).

Knowledge Management at Ernst & Young

Ernst & Young is one of the "Big Six" professional services firms which traditionally offered audit, tax and management consulting. In 1993, the operational vision for the consulting practice involved five key processes: sales, service delivery, people and knowledge. By 1995, the knowledge strategy had been formalized into an approach entitled *Accelerated Solutions Environment* which involved the rapid application of knowledge, models and approaches to client situations.

The firm created a position of Chief Knowledge Officer with the following responsibilities: oversight of processes and technologies related to knowledge. Also, several high level committees were created to examine the direction of knowledge management within the firm.

Due to the scope and geographical distribution of the Ernst & Young knowledge base and its users, technology had to be used as an enabler for its knowledge workers. Lotus Notes was selected as the primary technological platform for capturing and disseminating internal knowledge. Thousands of databases were in use for network and focus groups, as well as, key documents. In 1996, exploration began for use of a Web based intranet. Infrastructure investment occurred which benefited knowledge management and technology (Davenport, 1997).

Microsoft Knowledge Management Case Study

Microsoft's IT group has invested time and resources into identifying and maintaining knowledge competencies. Microsoft hired a program manager to take on the issue of knowledge competencies. The manager's goal was to create an on-line competency profile for jobs and employees. The project was called "Skills Planning Development" or "SPUD". SPUD is not focused on entry level competencies, but rather competencies needed to stay on the leading edge of the workplace. The five major components of the project were:

- Development of a structure of competency types and levels.
- Defining the competencies required for particular jobs
- Rating the performance of individual employees in particular jobs based on competencies.
- Implementing the competencies in an online system.

- Linkage of the competency model to learning offerings.

SPUD involved building an on-line system that contained the competency structure, the job rating system and ratings database, and the competency levels for employees. The system had a Web front end for easy access through Microsoft's intranet.

Implementation of these knowledge management strategies proceeded on an international level involving a large number of employees and utilizing a cross-section of all job types. (Davenport, 1997).

Knowledge Management at Hewlett Packard

In mid-1995 there were several knowledge management initiatives at Hewlett Packard. A series of workshops on the topic were facilitated by the senior leadership of the firm. The goal of the workshops was to bring together a diverse group of people who were already sharing knowledge or were interested in the topic. The workshop objectives were:

- Sharing of knowledge through informal networking
- Establish a common language and management framework for knowledge management

As a result of the workshops, HP discovered twenty sites for knowledge sharing. One example was a training database ; an on-line knowledge database that contained training issues, topics and techniques. Other Hewlett Packard knowledge sharing operations included a guide to human knowledge within HP's laboratories. The goal of this database was to identify expert profiles, or guides to the backgrounds and expertise of individuals who were knowledgeable on particular topics. Hewlett Packard Product Processes organization developed three knowledge sharing products: competition information, research information, and marketing intelligence (Davenport, 1997).

CONCLUSIONS and RECOMMENDATIONS

The results of the survey indicate an integral part of enterprise activity is knowledge management. From massive intranet based repositories to corporate-wide smart document applications, a constantly changing environment seeks to optimize knowledge management practices and leverage intellectual capital.

In each of the case studies, it is clear that if knowledge is to be leveraged, it must first be categorized. The thesaurus-based, matrixed approach used by Teltech may be the "best practice" since knowledge is usually communicated and sought in words.

Substantial progress has been made in Ernst & Young's approaches to knowledge management, but several challenges remain, inclusive of: embedding knowledge in an ever changing technological environment; maintaining support with increased usage; and buy-in from front line staff.

Also, the task of measurement of return on investment is difficult at best. A 'dashboard' was created assessing such topics as value delivered, reusable content created, thought leadership, presence of subject matter expertise and state of networking environment. It was noted that "some level of faith" was required to fully justify knowledge management investments. The revenues in the US consulting practice in 1995 rose by 44% which surpassed all targets ((Davenport, 1997).

The survey reflected a small percentage of industries utilized ROI as a measurement tool. Creative measurement tools are being utilized such as access by line, employee turnover, process cost, performance improvement and product/service innovation.

Standards bodies, ranging from the International Accounting Standards Committee to The Conference Board are making important progress on performance measurement methodologies (Petch, 1999). Performance measurement will be a key issue in knowledge management initiatives since there is little precedent upon which to establish ROI. Balanced scorecard measures inclusive of financial, customer, internal processes, and innovation have been proposed for knowledge capital (Kaplan R., et al, 1996).

As an emerging and dynamic discipline, the creation of a standard measurement of knowledge management reflected on the balance sheets, is still in the formation stage. Once achieved, the result will be a rapid response from global business leaders, to implement knowledge management "best practices" in order to remain competitive.

In the case of Microsoft, it is evident that for a project to be successful it would require the involvement and teamwork by everyone in the organization. This is a model which is clearly valid in business practices, but essential to the discipline of knowledge management. The Microsoft case study reveals security issues pertaining to building a competency database which included access rights based on the nature of the data; "people data".

The survey results showed 68% of respondents utilized networks as a technological tool for implementation of knowledge management practices. The continual development of adequate security for knowledge-based network applications is critical to this emerging discipline. In order to effectively leverage intellectual capital through technology based solutions, the issue of security must primarily be addressed.

Hewlett Packard has recognized that there are several knowledge sharing initiatives, such as training techniques, marketing intelligence, etc.. which can advance their position in a global marketplace.. It is clear that the level of resource commitment in gathering data and dissemination of data is high, but the gain is important to remain competitive in a global marketplace.

Analysis of these conclusions, implies that important aspects of this discipline are regarded in similar manner throughout the intellectual and business community. These aspects include a direct correlation between the implementation of knowledge management “best practices” and quality improvement, as well as, improved productivity.

Hedlund and Nonaka reinforce the important idea that not only the success, but also, the very survival of organizations will depend, in large part, on how well they create, transfer and exploit their knowledge resources (Hedlund, G. et al, 1993).

Knowledge management implores you to look at informal networks and protocols, any and all approaches to sharing experiences and know-how, as well as any and all cultural, technological and personal elements that spur creativity and innovation in response to changing stimuli. The effects of knowledge management on quality and productivity are evident by an internal and external awareness of collective strength and the ability to respond and instantly organize to meet market demands and opportunities. The competitive advantage for the new millennium is knowledge management.

Literature Review

Arthur, B. (1994), *Increasing Returns and Path Dependence in the Economy*, University of Michigan Press, Ann Arbor.

Davenport, T., DeLong, D. and Beers, M. (1998) “Successful knowledge management projects”, *Sloan Management Review*, Vol 39 No 2, pp. 43-57.

Davenport, T. (1997) “Knowledge management case studies”, Graduate School of Business, University of Texas at Austin URL: www.bus.utexas.edu/kman.htm

Earl, M. (1994) “Knowledge as strategy: reflections on Skandia International and Shorko films”, *Strategic Information Systems: A European Perspective*, John Wiley & Sons, Chichester.

Edvisson, L. and Malone, M. (1997), *Intellectual Capital*, Harper, New York.

Grant, R. (1996) “Toward a knowledge-based theory of the firm”, *Strategic Management Journal* Vol 17, pp. 109-122.

Hedlung, G. and Nonaka, I. (1993), "Models of knowledge management in the West and Japan", *In Implementing Strategic Process: Change, Learning and Cooperation*, Blackwell, Oxford.

Hibbard, J. (1999) "Protecting innovation", *Information Week*, Issue 722, pp. 51-65.

Hine, M. and Goul, M. (1998) "The design, development, and validation of a knowledge-based organizational learning support system", *Journal of Management Information Systems*, Vol 15 No 2, pp. 119-152.

Huber, G. (1991) "Organizational learning: the contributing processes and the literatures", *Organization Science*, Vol 2 No 1, pp. 88-115.

Kaplan, R. and Norton, D. (1996) *The Balanced Scorecard*, Harvard Business School Press, Boston.

Leonard-Barton, D. (1995), *Wellsprings of Knowledge*, Harvard Business School Press, Boston.

Malhotra, Y. (1998) "Deciphering the knowledge management hype", *Journal for Quality & Participation*, Vol 21 No 4, pp. 58-60.

Manasco, B. (1996) "Leading firms develop knowledge strategies", *Knowledge Inc*, Vol 1 No 6, pp. 26-29.

Mantelman, L. (1999) "Technologies of knowledge management", *Knowledge Management*, Vol 2 No 2, p. 67.

Nonaka, I. (1991) "The knowledge-creating company", *Harvard Business Review*, Vol 6 No 8, pp. 96-104.

Petch, G. (1999) "Memo to executives: get out of swann's way", *Knowledge Management*, Vol 2 No 3, p.16.

Seely-Brown, J. and Duguid, P. (1991) "Organizational learning and communities of practice: toward a unified view of working, learning and innovation", *Organization Science*, Vol 2 No 1, pp. 40-57.

Senge, P. (1994), *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization*, Doubleday, New York.

Smalley Bowen, T. and Scannell, E. (1999) "Making knowledge mean business", *Info World*, Vol 21 No 8, pp. 69-70.

Stewart, T. (1997), *Intellectual Capital: The New Wealth of Organizations*, Doubleday, New York.

Sveiby, K. (1997), *The New Organizational Wealth: Managing and Measuring Knowledge-Based Assets*, Berrett Koehler, San Francisco.

Archiving Electronic Grey Literature of Korean Academic Societies in Science and Technology

Cho, Hyun-Yang ; Choi, Seon-Heui ; Choi, Sung-Pil ; Shim, Hyung-Sup
Korea Research and Development Information Center (KORDIC)

Abstract

Korean academic societies in the field of science and technology are the major sources for producing STI (scientific and technological information) in Korea. They produce various kinds of printed materials including journals, proceedings, preprints, seminar/workshop materials of the study group in a society, and so on. While journals and proceedings are published and distributed to members via conventional routes, however, other materials are not. The secretariat of societies can play important roles on collecting grey literature produced by supporting and controlling academic activities in each society. However, there are only few Korean academic societies to create databases by themselves on materials generated from academic activities.

KORDIC(Korea R&D Information Center) is responsible for archiving and distributing technical reports, generated by the result of government-supported researches. KORDIC also implements the project to support Korean academic societies to cultivate their own capability of creating databases and providing information services. 130 Korean academic societies in science and technology are participating at this project. The project consists of three different parts. The first part of the project is to create bibliographic and full-text databases on all the products of academic societies. The second part is to construct Internet homepage for each academic society. And the last is to assist to provide users with information service via Internet. KORDIC in lieu of attending academic societies collects all grey literature from members, constructs various types of databases on all the materials collected, and provides users with information services.

In this paper we will introduce the present status of archiving electronic grey literature generated by Korean academic societies and its services to users via Internet. We plan to create English version of electronic grey literature database produced from Korean academic societies. Although this project is initial stage, construction of databases on electronic grey literature and the study of efficient information services will be continued.

INTRODUCTION

Throughout all over the world, academic societies in the field of science and technology have been the important producer of current scientific and technological information that is produced by academic society members by national and international research activity and interchange.

Because of newest research developing related information of every field of researcher announced through research journals of relating society, proceeding, seminar, workshop as of first priority, we are in need of information service that user can reach out for the information that has been produced and circulated faster and easier.

In Korea, with exception of huge academic societies, most of academic societies are struggling to bringing it to life and barring it alive due to the lack of finance and uncertain establishment of gathering, manufacturing information, and service system, even though they realize the importance of information services.

As of today, in Korea, there are about 350 academic societies in science and technology field and due to division of studies and developing of new era, more of detailed new academic societies are appearing. At this time, about 10 Korean academic journals are being reported in SCI expended and also making a constant quality progress. This is the time that we are in need of constructing databases of all types of information that the societies have produced. To compromise this situation, KORDIC (Korea Research & Development Information Center) is starting society informatization project about academic societies in the field of science and technology in Korea. Society informatization project is to construct databases of information that each society produces, and to provide financial and technical supports for information service through an Internet homepage. In this study, we will discover the category and characteristics of societies' producing and circulating research information including grey literature and analyze the service KORDIC is proceeding and inquiring into more affective society's information service idea and system for later on.

WHAT DOES KORDIC DO?

KORDIC

KORDIC is a non-profit government research institute under the supervision of Prime minister of Korea that is built to suit the purpose of gathering and manufacturing the scientific and technological information and constructing it to databases and then provide service it through Internet. In Korea, non-profit sci-tech laboratories are established in

specialized fields, with the support from the Korean government, and as the laboratory that offers effective information service about all fields of science and technology KORDIC was established. Since 1993, after the establishment, with constant database construction we are managing 14 database service throughout several types of scientific and technological information and hold around 3 millions of data.

KORDIC's database structure forms mostly with bibliographic databases, facts and statistics databases, in depth and analytical information databases. Most of databases consist of Korea's own contents about information occurred in Korea.

KORDIC's bibliographic databases can be divide into two different types; one is the database that contains bibliographic information, such fields as title, author, descriptor, abstract, etc. The other is the database that includes both bibliographic and full-text information (image file or text file form). Informatization for Korean Academic Society is to meet the second example.

Informatization Project for Korean Academic Society

In Korea, the actual academic society informatization has started with manufacturing and distributing CD-ROM title. And lately, 3-4 years ago, since the Internet became an active method of academic information circulation we have launched informatization through Internet. Since the WWW which made Internet popular was develop among few physicists and biologists, also here in Korea, informatization has been tried in this field at the first time. At the present time, the some societies run their own information system. These are as follows; The Korean Chemical Society(<http://www.kcsnet.or.kr>)[1], The Korean physical Society (<http://mulli.kps.or.kr>)[2], The Biochemical Society of the Republic of Korea (<http://www.biochem.or.kr>)[3], Korean Mathematical Society (<http://www.kms.or.kr>)[4]

The Korean Chemical Society has been chosen by Korea Science Foundation as a trial society to support sci-tech society informatization running the KCSnet that serves through Internet. [5] The details of the informatization project are to build a homepage on Internet and construct various databases for information services to target users.

The academic society informatization has huge difference among societies. It differs from the knowledge and enthusiasm of the president in each society. And also, there has been a case that started with related foreign society's request.

Recently, 2-3 years ago, national organizations that support the academic society have set the informatization measure as the important evaluation measure, and relating foreign societies have asked the informatization measurement for International Corporation. These circumstances have come to maturity that the academic society has to be informationized. Most of academic societies in sci-tech field have been asked building a homepage on the Internet and creating a database on all types of information, produced by societies.

With these understanding, KORDIC 's society informatization area has been set, and main contents was set to build an Internet homepage that each society asks and to create bibliographic and full-text databases.

The Selection of Academic Societies for Informatization Project

The actual goal of society informatization project that KORDIC carries out is to build a information system, so that members of the all societies and information users can access the full-text database and use it in their own computer through the network. To do this possible we have decided to support the academic society with the respect to building the society academic information database, supporting for information service system, computing facilities, and cyberspace on the network. In the given space, each society builds its' homepage and manages it by itself. Data structure for the database of journal articles and current society's data runs after MARC, the standard format and supports basic computer and communication facilities. To make the goal come to life more effectively, we have set trial societies according to their capabilities of producing academic information. As of this time there are about 160 societies are participating.

Table 1 shows the number of participating societies and a representative society in different subject category.

<Table 1> Number of Participating Societies

Subject Category	Number of Participating Societies	Representative Society
Basic Sciences	31	Korea Environmental Sciences Society
Life Sciences	9	The Ecological Society of Korea
Agriculture	23	Korean Agriculture Society
Medicine/Pharmacy	13	The Pharmaceutical Society of Korea
Engineering	33	The Korean Society for Composite Materials
Machinery	9	Korea Society of Automotive Engineers
Electric/Electronics	6	The Korean Institute of Electrical Engineers
Information/Communication	8	Korean Information Science Society
Management Sciences	4	Korean Operations Research and Management Society
Home Economics	10	Korea Society of Costume
Others	16	Linguistic Society of Korea
Total	162	

We developed criteria to select qualified societies for informatization project since KORDIC plans to supply financial and technical supports to societies. Criteria to be selected are years of existence, publications, academic activities, subject category in science and technology, number of members, and whether or not registering at Korea Science Foundation, etc. Besides, we considered the size of the society since we believe a huge society in a specialized field might lead to medium and small size of societies in informatization. Criteria will be shown at Table 2.

<Table 2> Selection Criteria of Academic Societies for Informatization Project

Category	Details	Minimum Range	Remarks
Years of Existence		5 years or more	
Publication		At least 2/year	Academic journals, etc.
Academic Activities		At least 1/year	Conference, Workshops, Seminars to be held
Subject Covered		Science & Technology	Engineering and related areas
Member Size		500 or more	
Reputation			

SYSTEM DEVELOPMENT FOR ACADEMIC SOCIETIES

Information Resources

We have made a survey of information resources, which will be used to create databases for pre-selected trial society. We have identified the information resource of society by analyzing data that have been collected from 143 societies as of now and could divide into two different types of information, academic and non-academic. Information resource from an academic society can be categorized by journals, technical reports, conference proceedings, and workshop/seminar materials, etc.. There are also non-academic resources, such as announcement, member information, and so on.

Resources for academic information are magazines, journals, proceedings that has been published in society and materials from seminars, lectures, and workshops which hold academic value. These are the important resources for creating databases of informatization project.

Member information includes not only the private information of each member, but also the academic achievement, major area, present research field, so when someone looks for the professional of the field, can be utilized. And furthermore, this information with citation index on published articles of every member can later on be used in academic database project such as SCI.

Information on academic and non-academic events and data, sponsored by each society, such as all kinds of table, application form, article of association, members of the board, rules of submission of articles, and ceremony rules is also available. Among these, materials from conferences, academic seminars, lectures can be grouped as grey information source.

Collecting Academic Information

All the data produced by the societies are gathered in society office and being sent at KORDIC, based on the agreement between KORDIC and each society. The data that have been sent are all of information we have described above. Among these, substantial amounts are grey literature data. The data, the journal, magazine, proceedings, seminar materials, workshop materials that have been published in their society are collected in paper or digital format.

Creating Databases for Academic Societies in Science and Technology

Among the collected data, journal articles with academic value were built up with bibliographic databases and full-text image databases. Title, author, co-author, affiliation, source, publisher, publication date, Korean/English abstract of the thesis are included in the bibliographic database..

Full text image database was built up with mainly with tiff image file(CCITT G4 Fax form, 300dpi) and some of data was built up with PDF format.

Structuring Homepages for Individual Society

Homepage of the society was built up with constructing a database. In order to design a model for homepage, we collect examples of domestic and foreign society's homepage. After then, we designed basic forms for each society, and asked them to check if designed forms are acceptable. Then uniqueness and characteristics of each society was applied. On the society homepage, society main page, introduction of each society,

members of the board, society news, publication status, proceedings and journals published, related site to be linked was included. When the menu of homepage was structured with BBS, it was made easy to put on the information for the manager and also to the users as well. So it made to grow the homepage uses. Part of managing and renewing of the society homepage is handle by KORDIC, but when society appoints the web-master, he/she takes care of it.

Information service system via Internet of society's bibliographic information and full-text image database was connected with society's homepage by using information retrieval system, KRISTAL II of KORDIC so user can search on WWW and also have access to the full-text. Each society's own information was unified to Academic Society Home Village (<http://society.kordic.re.kr/>). This homepage will be in English also sooner or later, so foreign users can have access to it.

Information Services for Grey Literatures

In addition to bibliographic and full-text database service for journal articles, another major information services of this project is to provide users with full-text image for proceedings, materials of seminar or workshops held by each academic society. KORDIC builds up these materials into the database having same format with journal paper. Included fields for workshop materials are date, place, researcher, affiliation, subject area, and full contents.

Information services for each society include not only journal article service, but about grey literature service as well which includes full-text image service on proceedings, materials of each seminar or workshops. These grey literature structures a database with a same format of journal articles. This database was provided users with full-text services separately or offers all the grey literature of one society service. And at the same time, we put together the same subjects of the society and give service hole in one.

And also, journal article services are running in association with societies using homepage BBS.

And further more, the system for management of society members and management of society's journal article has been developed and has set the system to provide information service for journal articles and grey literature from societies on publication.

FUTURE PLANS AND CONCLUSION

Society informatization project has its own goal in developing the system for the society itself to provide services for the information that each society produces. KORDIC's role in this is to support each society to build the information service system by itself. And with the corporation of societies KORDIC offers the unified service to users so they can access various kinds of information on societies in one place.

Future plans are to make more of trial academic societies. KORDIC will keep on studying and developing more of changing information technology so the system that users can access high quality information much easier and collect the academic information, produced by the society, in full text as much as possible, so we can offer full text service in multimedia.

References

- [1] The Korean Chemical Society (<http://www.kcsnet.or.kr>)
- [2] The Korean physical Society (<http://mulli.kps.or.kr>)
- [3] The Biochemical Society of the Republic of Korea (<http://www.biochem.or.kr>)
- [4] Korean Mathematical Society (<http://www.kms.or.kr>)
- [5] Choi, Seon-Heui and Hyun-Yang Cho. (1988) "A Study on the Information System and Information Resources from Korean Academic societies in Science and Technology, *KOSTI '98 Proceedings*, pp.231-236

Mine Maps as Grey Literature

By Linda R. Musser,

Head of the Earth and Mineral Sciences Library,

The Pennsylvania State University, 105 Deike Building,

University Park PA 16802; 814-863-7073; Lrm4@psu.edu.

Maps are a tough format - frequently big in size, awkward to store, and of little or no interest to many librarians. Even among the map library community, mine maps are the ugly duckling of the map library world. A narrowly focused, unglamorous, rarely collected stepchild of the cartographic community. Yet mine maps are repositories of unique, difficult to replicate information illustrating location and extent of historic and current mining activities.

The majority of maps are governmentally produced, a trait that qualifies most maps for grey literature status. Fortunately, such are usually published in significant enough quantities that they can be collected and held in multiple repositories. Their content is frequently of broad enough interest that many map libraries collect, catalog, preserve and provide access to these materials. The same cannot be said for mine maps.

Firstly, these maps are not produced by the government in most cases. They are produced by the mine operator for internal use, although many governments require their creation and deposit with the state. Secondly, they are produced in single or very limited quantities, making them hard to collect. The subject matter is such that, to many librarians, it seems of limited interest or utility. This is far from the truth, of course.

Mine maps are of great utility to a variety of users. Homeowners (and realtors trying to sell homes) want to know if there is a mine underlying the property. Mining companies desire knowledge of the extent of previous mining operations in the event they choose to commence re-mining operations. Geologists utilize these maps to assist

in mineral reserve estimation, sample collection, exploration and mapping. Planners and engineers are interested from the point of view of project and community planning. Government inspectors require them for reasons relating to health and safety as well as environmental protection and redemption.

CHARACTERISTICS OF MINE MAPS

Mine maps, like all maps, come in a wide variety of sizes, colors and materials. Maps were frequently drawn on the most convenient and inexpensive material at hand. Thus mine maps are found drawn on newsprint, velum, linen, drafting paper, canvas, and mylar. Color may or may not be used depending on the mineral or mineral seams being represented in the map. For example, maps of anthracite mining operations commonly utilize color to indicate the extent of mining in a particular coal seam. Size is also tied to the type of material being mined. Again, using historic Pennsylvania coal mines as an example, anthracite mines tended to be limited in width but could be quite lengthy thus generating maps of reasonable width (137 cm.) but extreme length (366 cm.). Bituminous coal mines, on the other hand, were extensive in all dimensions requiring maps of equivalent proportions (e.g., 600 cm. x 900 cm.).

MINE MAP COLLECTIONS

As previously stated, mine maps are little collected by most libraries. Many of these maps are acquired by donation directly from the mining company as part of their archives. On occasion, mining engineers will also donate mine maps. It is quite rare, however, to find a mine map in a library outside a government agency. Given that many governments require that mine maps be deposited with the state, it is not surprising that the government is the most common collector of mine maps. Unfortunately, not all governments take good care of these unique records of industry. Nor are they uniformly concerned with providing access to these records. Conditions and access vary widely.

In the state of Pennsylvania, for example, mine maps are collected by no less than seven different agencies. Each agency has unique materials and separate inventory schemes. Additionally, the cooperation and communication among the agencies frequently leaves something to be desired. Also located in Pennsylvania is a national repository for abandoned mine maps. Operated by the U.S. Office of Surface Mining Reclamation and Enforcement, the National Mine Map Repository (NMMR) collects and creates microfilm copies of abandoned mine maps loaned to the NMMR. Its holdings, while extensive, represent less than fifty percent of the mine maps in existence in the United States.

CONCLUSION

Similar repositories exist in other countries as well and they face similar challenges of collecting, cataloging, preserving and hopefully providing access to these unique materials. It is hoped that librarians will become more aware of the value of these not-so-pretty stepchildren of the map library world and endeavor to do their part in preserving these unique records of industry.

BIBLIOGRAPHY

Musser, Linda R. and Lisa A. Wishard (1999), "Mine Map Repositories in Pennsylvania and the United States", *Proceedings of the Geoscience Information Society*, v.29, (in press).

Office of Surface Mining Reclamation and Enforcement, U.S. Department of the Interior (1997), *Mine Map Repositories: A Source of Mine Data*, U.S.G.P.O., Washington, DC.

Automatic keywording of High Energy Physics

David Dallman, Jean-Yves Le Meur

CERN, Geneva, Switzerland

5 October 1999

Bibliographic databases were developed from the traditional library card catalogue in order to enable users to access library documents via various types of bibliographic information, such as title, author, series or conference date. In addition these catalogues sometimes contained some form of indexation by subject, such as the Universal (or Dewey) Decimal Classification used for books. With the introduction of the eprint archives, set up by the High Energy Physics (HEP) Community in the early 90s, huge collections of documents in several fields have been made available on the World Wide Web. These developments however have not yet been followed up from a keywording point of view.

We will see in this paper how important it is to attribute keywords to all documents in the area of HEP Grey Literature. As libraries are facing a future with less and less manpower available and more and more documents, we will explore the possibility of being helped by automatic classification software. We will specifically mention a project being carried out at CERN (European Laboratory for Particle Physics) for testing this automatic keywording.

SEARCHING DOCUMENTS BY SUBJECT

There are two main uses of a bibliographic database. The first one is to search for a specific item which one already knows about, and wants to find out if the library has it, and if so, to get access to the document. This is the so-called referral approach, a bit like looking up a piece of information in an encyclopaedia. You know it is there, you just need the answer.

The other main use is when one has a specific problem in mind, and wants to find documents which address that problem. It is only with this second type of use that we are concerned here. Basically, this means a subject-based approach to the library collection.

1) Subject connections via references

There is already a system of searching academic literature in a thematic way without any kind of intermediate database. This is via the references to other work which have been an accepted and important part of scholarly publication since the very beginning.

Starting from a core document, one can gradually widen the scope using the references and hopefully arrive at some fairly complete set of relevant documents. The electronic age has again enhanced such an approach without however changing it in principle. References in an electronic document can be links to the electronic versions of the documents referred to.

The main obvious drawback in this approach is that authors may not have referred to all the relevant material, either due to deliberate omission or just because they do not know about it.

Another possible disadvantage is that by definition one can only refer to what already exists at the time of writing a document ! From the core document, past documents will be reached but all new documents will be missed. However,

this could theoretically be solved through a database of such references, by forward searching from a document to retrieve all other documents which have referred to it later.

In practice, the connections via references is not an adequate approach for users needing for an exhaustive list of available documents related to a given topic. It takes too long and the full coverage is not guaranteed. The other solution - querying directly a bibliographic database - is much faster but it may still result in an incomplete result.

2) Recall and precision in searching

In evaluating the value of search results, two different concepts are of great importance: the first is the recall factor and the second is the precision.

The recall factor measures how many of the total number of documents which should be retrieved by the search, are in fact retrieved. Some relevant documents may be missed by a given search strategy. Taken on its own, this factor should obviously be as close to 100% as possible.

$$\text{recall} = 100 * \text{number of documents retrieved} / \text{total number of relevant documents}$$

The precision measures what part of the documents retrieved actually belong to the desired sample, the rest being undesired documents which have somehow also managed to satisfy the search criteria. Again, when considered alone, this should also be as close to 100% as possible.

$$\text{precision} = 100 * \text{number of relevant documents retrieved} / \text{number of documents retrieved}$$

The problem is that these two measures of search efficiency are not independent, in fact there is an anti-correlation between them. If you try to get the recall factor as high as possible by using a more complex search strategy, you will also tend to pick up more "background" documents which you do not want. Conversely, if you want all your retrieved documents to be relevant, you will have to pay the price of missing quite a lot of relevant documents too.

Database search engines may offer some features which are designed to improve the precision of the search. Words can be strung together as a phrase, and this phrase searched for. However searching for a phrase of more than three words is likely to result in a low recall factor, because of the flexibility of natural language (particularly English) in representing nuances of meaning by variations in word order.

In another approach, limits can be placed on the maximum number of intervening words which are allowed to occur between a pair of chosen words (proximity searches). The CERN Library database has such a functionality but it is rather cumbersome to use at present.

Of course, it is not just the search strategy which counts, but the result of the search strategy when applied to the data. Therefore, exactly which data for a given document are available for searching influences the search result.

3) Data from the document itself which are available for searching

Here we are only really concerned with data concerning the subject matter of the document, so things like author names do not play a role. Of course, searching for an author can result in retrieving a certain subject, but it is almost never the case that this author is involved with all the documents in that subject area.

Traditionally, the title is the item of bibliographic information which expresses the subject content. However, a title is usually far too short to contain a complete description of the subject area in a way which can be used efficiently by a search engine. A specialist reading the title may understand what the document is about, but he is using all sorts of prior knowledge into which context he plugs the new title. Therefore the recall factor of a title-based search is likely to be low. Furthermore, as the number of documents in the database steadily increases with time, the precision of title-based searches is likely to decrease as well.

An extension of this, which has become much easier to realise for electronic documents, is that more of the text than just the title can be used for searching. In particular, extending the search to the abstract is a very useful step. This has been done with the CERN HEP database since we started handling electronic preprints in 1994. However it has to be remembered that all words in the text are treated equally, so a mention in an abstract of a term by way of contrast and not because it is dealt with in the document, will still cause it to be indexed.

In principle, the full text of the document could be used for searching, but in practice this has not been done for documents in the HEP field. Considering the huge number of documents produced in this field, searching on full text would probably give a good recall factor but the precision would be far too low to be really useful. This is surely why we do not know of any project for indexing the whole text of HEP literature.

An alternative approach is to supply additional data concerning the subject material, and to use this for searching.

WHY DO AUTOMATIC KEYWORDING ?

Adding of subject material is called subject indexation or keyword enhancement. When we say "keyword" it could of course be a phrase of two or more words. There are two very different ways of doing this : to choose terms from a fixed thesaurus or to use free keywords which can be chosen by the indexer at will. The strategy of assigning keywords will obviously depend on which parts of the document itself (title, abstract, full document) are also available for searching.

1) Adding data to the documents

a/ free keywords

Allocating keywords on a free basis could also use terms which are not present in the document, but in practice this technique is mainly used for adding useful words or phrases taken from the text, such as section headings and other specific words which could help in improving the recall factor of the search. Free keywords can also be useful for indexing terms containing special characters which would not be completely recognised if they appeared in the title or abstract. For example, the CERN Library database normally breaks off indexing a word when it meets a non-alphanumeric character in a title or abstract, but it can be directed not to do this for a keyword field. Thus particles called W⁺ and W⁻ would both be indexed as W in a title, but the full forms can be used as keywords and retrieved.

Free keywords can also be a useful way of adding synonyms of terms that appear in the text. But it would be better in general to handle synonyms at the search input end rather than adding them to each record when they occur.

b/ fixed thesaurus terms

The efficient allocation of keywords from a fixed thesaurus makes the most demands on the indexer, as the documents have to be well understood. The indexed terms may not appear in the same way in the text at all, which can give this method a big advantage over any strategy which just uses the text of the document.

Of course, such a method requires the existence of a complete, precise and up-to-date thesaurus, which is quite difficult to achieve in a rapidly-changing specialized research area like High Energy Physics.

2) Comparison between free keywords and fixed terms

In practice, these two forms of indexing are extreme cases. Real approaches have aspects of both, even though they may be closer to one than the other.

Thus, the drawback of having a fixed thesaurus is that the thesaurus itself has to be modified to keep up with developments in the field. This usually means that a new form of the thesaurus is issued at regular intervals, for example the DESY (Hamburg, Germany) HEP thesaurus has been updated every one or two years. Thus for searching

back over many years, each time period should in principle be combined with the relevant thesaurus terms for that period. In practice, this complex procedure is rarely undertaken by the searcher. It could be built in as a front end to the search, but this has not been done yet for any of the databases in our field which use fixed thesauri.

On the other hand, free keywording can be chosen to conform to a minimum set of rules, instead of being completely free and just taking the words as they appear. For example, it could be decided to choose singular forms instead of plurals. In fact, after a period of use, listing the terms which have been given as free keywords does give a sort of "thesaurus in practice", which can then be used to standardize the keywords which are subsequently assigned, in order to improve consistency.

3) Influence of the keywording on search quality

Including the abstract instead of just the title in searching (with no additional keywording) increases the recall factor but probably reduces the precision. Use of proximity searching could offset this loss in precision somewhat. It is straightforward to measure the change in precision (within a database which permits searching in title and abstract separately) but the absolute recall factor cannot be measured like this, as one has no way of knowing which relevant documents have not been retrieved at all !

The free keyword system is designed to be used in conjunction with the other data like title and abstract. If used with the title alone, it probably improves both recall and precision. But it does not give much improvement over using title plus abstract, as free keywords are most of the time words already present in either the title or the abstract.

The fixed thesaurus approach aims at describing each document by a series of thesaurus terms in such a way that both the precision and recall are 100%. This aim might not be achieved in practice if the expertise of the indexers leaves something to be desired. It is very important to realise that searches should only be made using the thesaurus terms assigned, all other text like title and abstract should be ignored. Some kind of measure of how much the thesaurus keywording improves search results can be obtained by searching for the particular thesaurus term in the title or in the title plus abstract.

The table below shows the numbers of documents found in some examples. For the chosen term in the DESY HEP Index (which covers published HEP literature), we look for the occurrence of this term as a keyword, then we look for its occurrence within titles in the same database. To compare with the abstract search, we use the CERN HEP database [1], whose coverage is similar.

We also look for the same term within the global scope of all eprints, published or not published. This gives an idea of the area which is not covered by the HEP Index.

Database:	DESY	DESY	CERN	DESY	CERN
Search performed	By Keyword in published HEP	By Title in published HEP	By Title or Abstract in published HEP	By Title in published and non published HEP	By Title or Abstract in published and non published HEP
Terms					
"Higgs Boson"	8	703	510	1034	1305
"Supergravity"	5945	1791	1000	2714	2908
"Duality"	6257	1073	779	1557	2591
"Interface"	409	95	272	422	1005
"Bifurcation"	57	41	80	80	287
"Dielectric"	293	111	108	240	450
"Graphics"	147	9	41	140	161

"Measure"	364	273	582	607	2556
"Pair"	228	1279	1000	2151	3804

Comparison of searches by thesaurus terms and searches by title/abstract in the DESY and CERN databases (28/09/1999).

The differences in the results are striking. The most reliable numbers in terms of precision and recall are the ones in the first column.

This means that when a user finds "Higgs Boson" in 703 titles of published HEP literature, only 8 of them only are really relevant to this topic. On the other hand, while a user may find the word "supergravity" 2714 times in the title, or 2908 times in the abstract of grey or published HEP literature, the number of documents actually relevant to this subject is more than double this. The other examples show the same kind of mismatch.

Moreover, it appears that the quantity of HEP literature without a classification (because it is not published) is quite large.

4) Conclusion

The added value of keywording based on a thesaurus is obvious, even when many other bibliographic fields are searchable. There is a direct relationship between the added value of keywording and the number of searchable documents: the more documents you keep, the more you need keywording.

A simple subject allocation cannot be satisfactory in the long term. Subjects need to be refined till they actually reach the precision of a thesaurus. The permanent increase of papers available in HEP will lead to a chaotic situation for Information Retrieval if a complete effective classification is not undertaken. In the next section, we will see what has been done so far in the High Energy Physics area.

Indexing by subject specialists is by the far the most precise method, but it is costly in terms of time and it requires highly-qualified people to do it. The question arises as to whether one could achieve a useful result by some automatic procedure based on the text of the title, the abstract or the full document.

TOWARDS AUTOMATION IN HEP

Before considering the automation itself, we give an overview of existing classifications in HEP. We describe the HEP specificity regarding the development of a keyword assigning expert system. Finally, we explain the tests that are currently being carried out at CERN.

1) Existing Classifications in High Energy Physics

Manual keywording has been carried out at DESY for more than 30 years. It covers all published articles in the various areas of HEP. The DESY HEP Index publication was the main output of this activity from 1963 to 1997. This publication itself then stopped but keywords are still allocated and they are searchable on the Web interfaces of the DESY [2] and SLAC [3] (Stanford, California) library catalogues.

A manual keywording activity used to be done at CERN as well. It started in 1983 with free keywording and was based on the HEP Index thesaurus (from 1989 to 1992). After this it was stopped due to lack of manpower.

Examples of fixed "commercial" thesauri are those used by INIS [4] (International Nuclear Information System, Vienna) and INSPEC [5] (Physics, Computing and Electrical Engineering Abstracts, UK). They are built manually and access is not free of charge. They are not sufficiently specialized in the HEP area and so are not really adequate for dealing with HEP literature.

Today, some articles do contain subject information supplied directly by the authors (usually only when the journal makes it a condition of publication!). So some journals have keywords, and quite a few journals have adopted the

Physics and Astronomy Classification Scheme (PACS) classification [6] supported by the American Physical Society. However, these approaches are far from being complete, so they are not useful for global searching. Also the PACS classification is still too broad for detailed searching in a narrow field such as particle physics.

On the contrary, in the case of books, where the Universal or Dewey Decimal Classifications are widely used, this approach can be very useful for retrieving all books dealing with a particular subject. A Web interface enabling searchers to browse a partial UDC index exists for CERN Library book catalogue [7], HEP preprints and published articles have no such world-wide recognised classification.

The CERN project, in collaboration with DESY and SLAC, is to use an expert system for automatically deriving keywords and then to map them onto the DESY HEP Index. In other fields such projects have already been carried out rather successfully. The Medical National Center (MNC) [8] and NASA [9] use machine-aided indexing for example, to speed up their classification.

2) Particularities of HEP literature

Natural language contains a huge vocabulary and the syntax of languages is very complex. In traditional literature, a text can be processed by considering words as individual items. A dictionary of single words can be used as the basis for creating a knowledge base. In scientific literature, we consider that the meaning is expressed mainly through multi-word terms ("noun phrases").

In HEP, documents contain many particle symbols or equations which may be among the most relevant noun-phrases in the document. Describing the syntax of the sentences present in HEP literature requires at least the definition of a new type of word: the particle symbol.

In addition, the knowledge base needs to be set up differently for experimental and theoretical documents. It is also planned to handle another specific dictionary for technology-related papers.

Another particularity is the size of its electronic grey literature. It amounts to more than 100 000 documents since 1994 and is growing at the rate of about 20 000 per year.

3) Sokrates Learning System

SOKRATES [10] stands for "Self-organizing Object-oriented Keyterm Recognition And Text-Editing System".

It derives from natural language key terms and keywords. Each new piece of information treated by the system is used to update a knowledge base. A learning system like Sokrates can be compared to a compiler: the input is a text written with a known syntax. The output is a condensed executable, like the set of key terms.

In terms of the earlier discussion, the Sokrates approach belongs to the free keywording type, where the free keywords must appear in the text and cannot be invented (except perhaps for synonyms if one chooses to build them in). So, for example, there is no way that this algorithm can return the term "Kac-Moody algebra" when the abstract says "graded Lie algebra", even though these are two names for the same thing.

The test of the software is divided into two parts: the derivation of the best key terms and their mapping onto the thesaurus.

a/ The Term Derivation

To run the extraction of the Key terms, three basic components are defined:

- ✓ A complete dictionary which is created and continuously updated. In this dictionary, individual words (any possible character type) are kept with the following two main attributes:
 - a code or type of word: "General", "Left", "Right", "Stop-word", "Particle", etc.
 - its frequency: number of times the word has been encountered in all documents processed so far.

- ✓ A knowledge base which stores all the key terms (single or multiple words) which have been selected together with their frequency.
- ✓ The rules for describing key terms (the "Text Description Language") which are expressed using the type of words. An example of a rule is:
L A P G R ... (Left, Stop Word, Particle, General, Right...), where L (R) enables one to specify that a word would only be significant when it appears to the left (right) of another relevant word.

An inference engine is able to match any rule to any text using the above rules and the dictionary of single words.

When dealing with a new document, Sokrates extracts all individual words and distinguishes old and new words. It counts repetitions and updates the dictionary. For new words (if any), it can ask an operator to provide the word characteristics (code).

In the next step, the term selector selects candidate terms. It uses the dictionary and the inference engine to extract all possible noun phrases. After this pass, a set of "valid" and "garbage" terms are available. The selector compares these derived terms to the established knowledge base, and keeps the ones for which a frequency threshold has been reached. The threshold can be defined differently according to the number of words in the noun phrases. For example, we could require that single words must have a frequency of 10, two-words terms a frequency of 5, etc.

If the number of derived keywords is too small, a third pass of the selector can be undertaken, with a reduced threshold definition.

b/ The Thesaurus Term Mapping

Different situations may occur:

- the key term exists in the thesaurus: the mapping is straightforward.
- the key term is similar to a thesaurus term: the correspondence can easily be established.
- the key term does not exist in the thesaurus: a subject indexer could associate one or more terms from the thesaurus to each key term found. This would only need to be done once for the whole dictionary; it would remain valid for all incoming documents. Only new key terms would need to be associated at later stage. Thus the savings over indexing on a document-by-document basis could be considerable.

4) CERN test: the status

The data for the test consists of about 1400 abstracts of preprints in each of the three fields of experimental high-energy physics, theoretical high-energy physics, and technological articles relevant for technology transfer. These have all been keyworded using the manual method.

The object of the test is to compare the automatic procedure with the manual method of keywording in use at DESY. We intend to find out whether the automatic procedure can be tuned to deliver keywords of a similar quality as DESY does with the manual method. Even if this turns out to be unrealistic, the expert system could be used as a Machine Aided Indexing (MAI) system to propose keywords to the indexer.

When the program has been tuned on these samples, it will be supplied with other samples of the same size. In comparison with the manually extracted keywords, the keywords extracted automatically should be as similar as possible in quantity (number per abstract) and in character.

70 000 words have been used as the "learning text" so far. Among the last 4000 words processed, only 200 words were unknown to the existing dictionary and required an input from an operator.

250 rules have been defined. Any sentence, parsed through these rules, will end up with one or many possible noun phrases.

The thresholds are still being defined: they need to be regularly adjusted as more documents are processed.

CONCLUSION

We can draw three main conclusions from the analysis and tests done so far:

- 1) The necessity of automatically keywording High Energy Physics Grey Literature is obvious.
- 2) We are optimistic that we will be able to build a valid knowledge base of noun-phrases, using Sokrates.
- 3) It is not yet clear how difficult it will be to successfully map this base onto the HEP Index thesaurus.

In all cases (whether the mapping is robust or not), the idea is that whenever a new document (with an abstract) is entered into the system, the expert system quickly delivers a set of key terms. This output can be added to the database straight away or it can be mapped to the thesaurus to try to deliver an "assigned term" and finally, it can be checked by experts before being loaded.

If the CERN test is successful, it will be run on a large scale in order to progressively cover all HEP preprints not yet classified. In addition to the traditional keyword searching option, a new utility will need to be developed to offer searchers a simple way of browsing through the thesaurus.

REFERENCES

- [1] <http://weblib.cern.ch/Preprints>
- [2] <http://www.desy.de/library/homepage.html>
- [3] <http://www.slac.stanford.edu/spires/hep/>
- [4] <http://www.iaea.or.at/programmes/inis/inis.htm>
- [5] <http://www.iee.org.uk/publish/inspec>
- [6] <http://aip.org/pubservs/pacs.html>
- [7] <http://weblib.cern.ch/share/udctree/udctree.html>
- [8] Wright, L.W., Grossetta Nardini H.K., Aronson, A.R., Rindflesch, T.C (1999) " Hierarchical Concept Indexing of Full-Text Documents in the Unified Medical Language System Information Sources Map", *Journal of the American Society for Information Science*, vol 50 No 6, pp. 514-523
- [9] Silvester, J.P. (1997) "Computer Supported Indexing: a History and Evaluation of NASA's MAI System", *Encyclopedia of Library and Information Science*, vol 61 Jan 1998, pp. 76-90
- [10] Sokrates Learning System, Dr. I. Steinacker, Intelligent Text Analysis and Management, Jenewein Burotechnik Innsbruck Austria.

Trends in Publishing Academic Grey Literature: Examples from Economics

Fethy Mili

Librarian, Département de sciences économiques
Université de Montréal
6128, Succ. Centre-ville Montréal, Canada H3C 3J7
Phone 1 (514) 343-6111 - Fax 1 (514) 343-7221
Email: fethy.mili@umontreal.ca

Abstract

What impact might electronic publishing have on grey literature? This work tries to give some answers based on changes in the publishing of economics research preprints or working-paper series.

Institutions and some members of the economics community have created archives to facilitate access to working paper series. In this document, we evaluate the prospects for electronic publishing of economics preprints using an inventory of material available from institutions' web sites. Also the paper contains an overview of some working-paper archives managed by economists.

1) Introduction

This paper deals with preprints in economics. Exchanging preprints or "working-paper series" as they are commonly known among economists had, been the traditional way for informal communication of recent research findings. Electronic dissemination of recent scholarly works has grown with the use of the Internet and the World Wide Web (WWW) facilities. Indeed, it is well documented now that computer and electronic dissemination of knowledge have changed the way of life within the information world. The incursion of the Internet as a gateway to the global information environment plays a significant role as a support in daily research activities.

Within universities, the Internet has become one of the major information vehicles particularly among the scientific community . It is quite easy to search information from the Internet for processing and manipulation. Indeed "the element of scholarly research most affected by the Internet is that of communication at all stages of the project, from study design and data gathering to project administration and dissemination of outcomes". P. Warren-Wenk and Vivienne Monty (1996).

For economists, as well as for many other academics, "Scholarly communication in the sciences holds, at its cornerstone, the tenets of publishing quickly the results of scientific inquiry or findings, and distributing the results of research expeditiously to the academic and industrial communities" as mentioned by J.Gelfand (1998). This situation has increased interests to find out quick channels to communicate research findings among colleagues and common academic public.

The delay associated with publishing in commercial journals, promotes the importance of working-papers series among economists. These documents were an assigned candidate for electronic publishing, distribution, and sharing via e-mail services and Websites. Robert P. Parks as pioneer of promoting the use of electronic infrastructure for economists describes with evident enthusiasm the success of E-Print Archives at Los Alamos National Laboratories which provide readers with easy access to preprints. This philosophy was the corner stone in creating the pattern used in managing EconWPA (<http://econwpa.wustl.edu/wpawelcome.htm>)

Since the beginning of the 1990s economists launched several on-line working-paper networks in order to intensify awareness among the economics community about recent findings. According to S. Karlsson and T. Krichel, the electronic working-paper series in economics appeared in 1993 and the production has grown as the use of the Internet features. Their paper pointed out the progress realised in electronic publishing of working-papers during the last few years. Indeed, many members of the economics community rely on the Internet to develop a self-publishing procedures. P. Warren-Wenk and V. Monty (1996) suggested that Web features provide faculty with a friendly environment in order to establish their own homepages and disseminate their writings.

Others projects have been created in order to promote electronic publishing by building either central or shared archives with free access. These projects aim to enhance the speed of access to publications. Economists were invited to participate by submitting their prepared papers to the archives in order to facilitate access to preprints. Whether they are centralised or decentralised, these archives aim to be a reference source to provide information about preprints in economics and to avoid the evident chaos (L.A. Davidson and Kimberlet Douglas, 1998) as well as "a growing disenchantment" within the Internet as described by A. Eisenberg (1997). However according to W.L. Goffe and R.P. Parks (1997), "Authors who want their work to be read need to post their papers to a working-paper archive like EconWPA, or register the paper with an index like WoPEc".

The creation of such archives (NetEc, EconWPA, SSRN/ERN ...) has been associated with the arrival of local archives among academic and research institutions which have started publishing working-paper series in electronic format as "it is assumed that patrons would like the ability to have electronic access to serials either from their desktop or from a convenient location." as noted by Tenner, E. and al. (1998). For many academic institutions and research units, announcements of preprints in economics have become a part of their websites in order to inform Internet users about the research activities taken by faculty and other researchers. Also advertising research findings is a way to promote the dynamic work of research teams within the institution. Currently there are several hundred universities and research units with Web and FTP servers that provide online information about working-paper series.

In this paper we will try to sketch the changes which have occurred in publishing and distributing paper series in economics. After several years of introducing electronic publishing, what does the world of grey literature in economics, especially preprints look like? In other words what has happened to the access to and availability of paper series in economics since institutions started to be interested in electronic publishing. Is it more easy to get a reference about working-papers? Do the announcements provided by the websites supply users with links to online papers? In short, has the incursion of the Internet into the research environment increases the possibilities for user to get working-papers and bibliographic information?

Meanwhile, the arrival of electronic full texts has not decreased the publication of hard copies. Indeed our daily experience shows that the production and the distribution of printed working-

paper series is still prosperous and only some academic institutions have thrown aside production in printed form while offering only electronic version of these papers. One might be surprised to realise that so many organisations still rely on the “conventional” method to advertise research results despite the promises of the electronic way of publishing.

In other words the use of the Internet to disseminate preprints in economics has multiple faces and we shall test the publishing services provided by the institutions in order to understand the trends in the availability of electronic full texts of working-paper series. This may help us as librarians to deal better with a changing world of information.

Like other grey products, investigating the situation of preprint literature in economics is a challenge because of the lack of a centralised project system for storing and distributing papers. As a result, it is difficult either to get or to define a clear view of what is effectively produced by different economic research units. In fact one can come across downloadable papers in homepages of authors as well as in other sites managed by academic departments and research centers. As this paper is concerned with the preprints prepared by the institutions (department of economics and research institutes ...) we don't attempt to evaluate the individual contributions widespread through the Internet. Also, we will not use the contents of particular projects like WoPEc. The reason is that this kind of archive relies basically on volunteer contributions and we believe that a reliable way to measure the changes is to see the institutions' participation to provide digital versions of working-paper series.

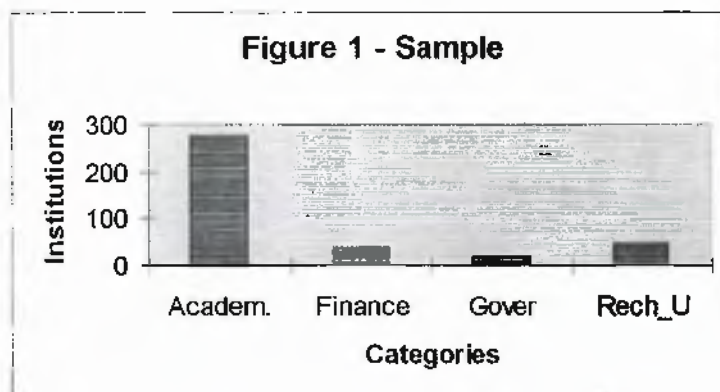
For this study we will use the series we receive on regular basis at our library. Our collection of working-papers counts more than 410 series. In 1993, these series were used to launch NetEc/BibEc with is a large bibliographic database of working-paper series. At present BibEc holds 50,000 bibliographic records of working-papers series. When created, the service included bibliographic records on some 35 000 economics working-papers (W.L. Goffe, 1994).

2) Data Description

The data we employed result from a first-hand compilation of elements related to working-paper management. However we are confident that the sample used would give a valuable appreciation of both past and future changes in electronic publishing of preprints in economics. Thus a sample of 383 Websites was used to gather information about the state of publishing

paper series via the Internet. We didn't contact providers and working-paper coordinators in order to supply details about the availability of their publications on the Internet. The reason was to simulate a searching process that could be made by a common user and to find out what kind of information he can get.

It is clear that we did not rely on any special scientific method of sampling and the sites we've examined were selected randomly. The series were drawn from our local collection as our library has a strong tradition in collecting and processing working-paper series in economics.

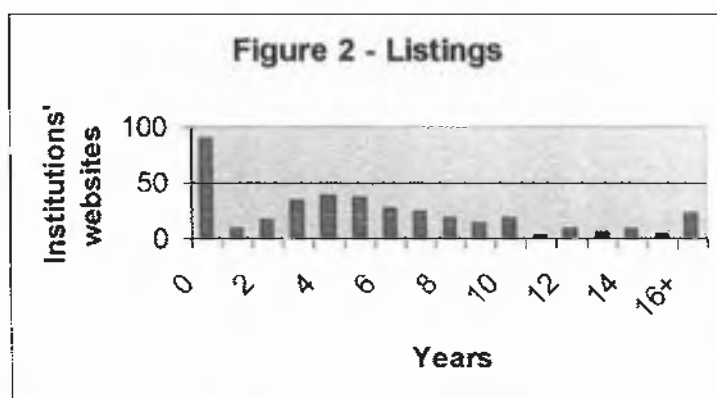


All the sites visited belong to departments of economics and research units. Figure 1 provides some details about the series sources. As shown, academic departments of economics present the major part of the sample. Indeed 276 websites checked belong to universities while the other providers are related to national and international financial institutions as well as to government agencies and to independent research units. As already established, academia is an important producer of knowledge; and, an important part of grey literature in economics is still produced by and for academics.

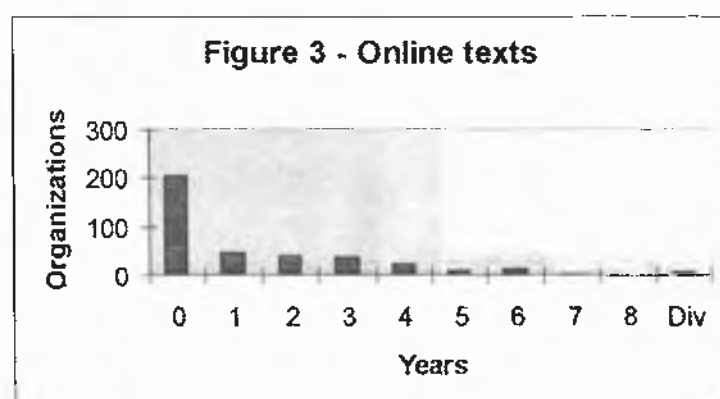
3)Result

Three different approaches have been observed while monitoring of the different websites:

3.1- Institutions that provide bibliographic lists of working-papers series produced by their faculty and searchers: These lists describe the contents of the series. An abstract is always given with each record at least for the recent years. These lists can be very helpful for people who need to be informed about the last preprints produced and in many cases the documents could be ordered either from the institution or from the author. In some cases a working-paper coordinator is associated to the Website, this makes it easy to get a paper when desired. We've also noticed that the great part of the sites visited have listings for 3 to 7 years. Only few institutions provide lists since more that 8 years, as Figure 2 shown



3.2- Institutions that make available access to the full text of papers: in that case, one can find information about how download the file as well as some details about the format in which the files are stored. Online files (downloadable full text) are less available than one might expect. Our inventory shows that only 177 institutions (around 46% of the sample) are involved in offering online files of the paper series produced. The other indication is that 146 organizations (more than 80%) provide electronic version of the series for 1 to 4 years. See Figure 3.



3.3 Institutions without information about research preprints: these organizations are not persuaded that scholarly works could be well disseminated via the Internet. It is an important "indifferent group" (E. Shreeves, 1992) which is not committed or convinced about the usefulness of academics going digital in order to increase the communication of their findings.

Also the inventory permitted us to evaluate the updating frequency of the information provided by the institutions websites. In fact, one of the features that makes the information valuable is to keep it up-to-date. In reference to the listings available, 199 out of 293 contain data about publications produced in 1999 when 108 among 177 providers offering downloadable full text papers hold material produced in 1999. In addition, a link seems to exist between the starting of making the material available (lists & online files) via the Internet and its up-to-dating. Indeed institutions that keep the sites up-to-dated have mostly began an online services during the last 5 years.

In order to summarize the results of the inventory one can raise the following points:

- 1) Information about working-papers in economics is more and more available. It is the most positive point for enhancing awareness about economics' grey literature. Information professionals can be better informed about grey literature production. Indeed, besides its quality, the information we found while browsing the website illustrates the impact Internet is playing to insure better advertising to grey economic literature.
- 2) Despite the increase in information about preprints in economics, the effective participation of institutions in starting and maintaining full text Websites is still conservative. Even if electronic publication of working-papers seems to be accepted by economists, the results suggest that organizations are not making the necessary effort to guarantee the availability of the papers as is the case for producing printed working-paper series.
- 3) The electronic availability of paper series is quite recent. It is an emerging service and we may expect an improvement of the organizations' participation in the future.
- 4) Up-dating information is an important element for accuracy of the material. Many providers don't add their recent publications to the website. This would reduce users' interest to rely on institutions' websites to stay informed. The facts suggest that the information professionals dealing with grey literature should keep on gathering material according to the traditional methods if they need to provide customers with recent publications.

4) Discussion

Beyond the deficiency of the present context there are facts, and one should realize the great work done by organizations in order to give “grey literature in sciences gains a more prominent role in a range of information use, access and dissemination activities” (J. Gelfand 1997). In economics, electronic publishing of paper series is in progress and while browsing the institutions’ websites we understand a real willingness among providers to disseminate information about the research preprints; however, one may require more commitment among organizations in order to obtain an effective fully electronic access to working-papers. Scholarly grey literature in economics is subject to an important transformation as it is taking advantage of computer-based networks and all the attendant technological features. As observed by William L. Goffe and Robert P. Parks (1997) “while some changes to an electronic world might be long in coming, and while there may be some confusion along the way, we expect that the ultimate outcome will be exciting”. Nevertheless at present, the eagerness to moving toward a digital future is not shared by all the community. Institutions seem to be unsure about how to proceed. As we’ve noticed, in some important organizations still stand aside from wave of change, mainly within a friendly environment created by software facilities for electronic publishing such as html and Postscript tools and pdf/Adobe Acrobat. Some common reasons could explain the reserve either to start or to improve electronic dissemination of research findings among some institutions:

- Technology;

Edward Shreeves (1992) observed that “The technology itself is in a state of constant flux with little probability for stability”. We believe that the context is not better today and “this forces one to realize that technology is never at a standstill and that the Web remains at a frontier stage, more embryonic than developmental” (J. Gelfand 1998). Changing standardization and publishing technology can be a source of hostility against technological features as well as increasing doubt about the Internet’s efficiency for communication.

- Commercial activity

Subscription to working-paper series is still frequently used if we need to get issues on a regular basis. Our experience confirms that we have always paid to receive paper series produced by some departments and by most major research institutions. Providing working-

papers series, through a subscription, has always prevailed among research institutions and the incursion of Internet and electronic publishing has not modified this traditional way of dissemination. In some cases subscription have become more expensive than before. Market laws don't incite these providers to change their behavior so they are not disposed to move towards digital publishing with free access.

- Cultural context

P. Warren-Wenk and Vivienne Monty (1996) have observed that "scholars seem to be using the Internet as a current awareness service". Indeed many faculty would appreciate using the Net to be entertained but they don't feel comfortable to be involved within an effective electronic environment. As noted by many, electronic publishing creates some ethics concerns mainly for scholars seeking academic recognition so they must publish their findings in a formal print journal. Uncertainty influences the institutions' commitment in to producing electronic versions of paper series on a regular basis. Peer reviews present an other great obstacle facing the acceptance of electronic publishing as a serious product. Many think that electronic publishing will be adopted by the scientific community only when it meets the needs of that community, solves problems related to print material and offers new opportunities that exceed the traditional way of disseminating knowledge. As long as these conditions are not satisfied, electronic publishing will remain a marginal medium associated to "professional pop art" (Raney R. Keith, 1998).

Some more prosaic reasons can also explain why some institutions delay starting electronic publishing of paper series is explained by Sune Karlsson and Thomas Krichel (1999) while talking about economists hesitation to participate to a central archives of electronic working-papers "Our theory is that economists have a built-in distrust of monopolies. In their book of tales, there are a numerous accounts of the welfare losses caused by monopoly supply. They may also be afraid of power accumulated by a person who controls a hard disk where the complete output of the discipline is stored."

All these points can justify the levels of institutions' engagement to introduce, and to maintain and to improve electronic publishing of preprints. Nevertheless, our hypothesis is that more involvement could happen within institution when people feel ready enough to start a new way of doing things. We believe that all organizations, mainly universities in North America have

the know-how in order to go on disseminating the preprint in electronic format via the Internet, but obviously print material seems to be enjoyed by many users. Also institutions have to be convinced that using cyber space is not a fashion or an “electronic gadgetry” as pointed out by Edward Shreeves, but an irreversible stream for information management and that “the benefits more than compensate for the effort” (S. Karlsson and T. Krichel, 1999). In between, information providers (universities and research units) as well as information professionals and clientele will keep on building a meaningful framework for the forthcoming electronic days. In any case the present hybrid context of producing working-paper series can't be harmful to grey literature which will continue to benefit from visionaires experiments as well as from the wisdom of its traditional managers.

5)Community projects

This paper is interested in the institutional contribution to working-paper electronic publishing. However we can't have a complete view of preprints services if we ignore scholars' participation in building and managing electronic papers projects. In this section we present a summary of two major initiatives conducted by some visionary economists in order to facilitate access to full text working-papers series.

It would be pretentious to fix a date for the starting of electronic publishing of paper series; however, it is admitted that during the early 1990s resources for economists were introduced to Internet. In 1994 William L. Goffe published an exhaustive list of material available on the Internet. Among the projects described in that article, NetEc and Working-Paper Archive (EconWPA) were announced as services dedicated to promoting the usefulness of electronic publishing of working-paper series.

These two projects are still leading to make economic grey literature as available as possible to the Internet users.

EconWPA (URL: <http://econwpa.wustl.edu/wpawelcome.htm>), is an electronic archive based at the Economics Department of Washington University. People can add their papers via a webinterface using any format. Working-paper are grouped in 26 subject areas, according to the classification scheme of the American Economic Association. Different methods of searching are available.

NetEc (URL: <http://netec.mcc.ac.uk/NetEc.html>) is an umbrella for a “group of projects to make the network more useful to the academic community in economics” (T. Krichel and T.

Wichmann, 1994). WoPEc is a service about electronic working-papers series. S.Karlsson and T. Krichel describe the service as a database focusing “on the collection of metadata about papers rather than the papers themselves”. WoPEc uses a standard protocol called RePEc in order to exchange data between participants without extensive coordination. This decentralized archives allow providers to contribute to WoPEc through the acceptance of a defined template format to display the information. According to coordinators, the project has enough success to be considered now as “the primary source about recent research in economics”. In fact, EconWPA and WoPEc are supplied by material from the institutions already producing electronic preprints and it would be difficult to evaluate the impact of these archives in order to convince more organizations without electronic projects to start new initiatives.

EconWPA and WoPEc have difference philosophies for managing paper series. The first site is claimed to be a central archive and its manager observes that “While economists are understandably skeptical about monopoly providers, a single database for each type of information may well be preferable to multiple,” (William L. Goffe and Robert P. Parks). On the other side, WoPEc stresses on the virtue of decentralized model. Currently, the two projects reflect the eagerness of a growing number of economists to participate effectively in the dissemination of research findings in economics. The projects maintain their activities and in both cases many efforts are used to keep the material available, accurate and easily accessible

6) Promises and Trends

6.1 Electronic vs. Printing

Over the past decade the face of grey literature in economics has changed profoundly. Preprints are much more available from different providers so students and faculty have virtually more opportunities to be informed about recent research. Print papers series are gradually being replaced by electronic publishing. Many organizations have shown a real willingness to move forward electronic publishing of working-paper series by cancelling the distribution of printed issues.

Nevertheless, one could ask whether the availability of the material via the Internet is a sufficient guarantee to ensure access to the publications. Our experience with working-paper

management advises us to be cautious. Indeed, the Internet free market cannot (at least presently) whiten the grey literature and make it more identifiable for the common user who still encounters important difficulties to get a valuable information through electronic media. Grey literature could be more spread through the network, however it can't be easily found. We still need human interfaces to gather the information because the Internet as a delivery medium creates a significant need among academics and other users to orchestrate the flows of information. It is quite important to recognize that publishing on the Internet "has created new models and opportunities, in an almost magical way" (J. Gelfand, 1998), but one has to make the difference between the hope to see the Internet solve some of the library daily problems related to grey literature (locating, retrieving ...), and the scholars who "are facing unprecedented information overload in their attempts to identify potentially relevant information sources" (G. McKiernan). As pointed out in W.L. Goffe and R.P. Parks (1997), "rational authors should realise quickly that readers will not find their work" unless it is available from a recognized provider. Since availability does not mean accessibility, one should continue gathering grey literature using the conventional pattern and using electronic material to complete users' needs. Many institutions are aware about the instability of electronic publishing on the Internet. So they keep on producing printed publications and we believe that this strategy will prevail for the long run nevertheless the pressure adopted by some "electronic-publication protagonists" (R. Heith Raney, 1998).

As shown in the inventory, electronic publishing of grey literature in economics is still a pilot venture and one has to wait in order to see whether incoming strategies which are being developed will influence the level of success or failure of publishing paper series. Sixteen years ago, J. Newcomb (1983) observed, in reference of electronic information future's that "the most obvious is that there is a lack of certainty as to precisely what the impact of the new technologies will be and how will they can be applied to publishing". As now, it obvious that technological tools increase the potential of electronic publishing but the chaotic reign within the Internet makes it difficult to get an effective use of the material available. So today's "most common complaints voiced by both experienced and new users of the WWW is its lack of structure and organisation" (David Revelt).

6.2 Standards

One of the major limiting factors to create a large acceptance of electronic paper series is the failure of agreement about a common way to advertise, to store and to display issues. While monitoring the different sites, we have encountered different presentation styles and storing formats. However some archiving projects such as WoPEc initiative, have established protocols that allow providers to exchange data. Currently, Postscript, PDF and HTML seem to be the most popular formats for publishing papers on the Internet because of their flexibility and the fact that many researchers can have access to the required technology to read the files. As observed by David Revelt (1998) "many sites have recognized that there is no common file format for electronic working-papers, and thus, that make no attempts to impose one". The Lack of standards could be a source of difficulties to access electronic paper series as users need to have the appropriate plug-in and viewers to display the files and many are convinced that "there is not, and most likely, will never be a standard format for working-papers or other Internet documents" (David Revelt, 1998). We believe that the format war will not take place despite of attempts by some archives managers to create a reference format to describe and to store electronic versions of working-papers. Standardization of electronic version for working-papers could not occur because paper series did not have a recognized model before introducing electronic publishing.

6.3 Providers

Our inventory related 3 major players within the world of electronic publishing of working-paper series: the institutions, the authors and some economics community members who have built special projects to coordinate the dissemination of paper series. We should note that each provider is motivated by a different interests in promoting the use of electronic material. The difference of opinions about going digital creates a additional obstacle to the acceptance of electronic publications by the scientific community. For organizations (Universities and research units), making available research findings on the Internet would be a signal for their commitment to innovate the procedures about advertising data about research related activities. Self-publishing is a result of the possibilities opened by the Web tools, so faculty can establish their own homepages in order to disseminate their works. Institutions and authors promote electronic publishing because of its optimal opportunity to enhance awareness about academic findings among the community and not to fight against traditional publishers. Indeed,

some economists involved in publishing projects consider electronic publishing on the Internet as an important step to breakdown the authority of the commercial publishers on academic knowledge. William L. Goffe and Robert P. Parks observed that "In academic, a primary goal is the growth, acquisition and dissemination of knowledge, which is aided by the freest possible access to information produced by and for academics" (1997). The electronic publishing and dissemination of research findings could help to keep off all the mediators.

6.4 Perspectives

The Internet state, the difficulties to use standards and providers different strategies confirm that electronic publishing still needs to define its contribution not as an alternative to printed material but as an effective tool to meet users' needs and to increase the effective dissemination of research findings. As argued by P. Warren-Wend and Vivienne Monty "electronic journal supplemented, but did not supplant print journal as a mechanism of monitoring developments in a field". In the same way, electronic working-paper series did not replace printed issues and our everyday experience attests that users are always seeking for hard publications for their research topics and we don't feel that some preprints on the Internet has changed neither the structure of our work, as librarians, nor the nature of the reference demands from the clientele. In any case, working-paper producers seem to continue publishing papers in printed form even if some organizations have initiated electronic projects. So the changes produced by electronic publishing on the conventional management of paper series are not so evident. As a librarian administering a large collection of working-papers, I am still discovering the ways to introduce Internet-based publications into ready reference services. In the short time, we believe that this part of grey literature in economics will continue to be produced in hybrid packages since the electronic world is a work-in-progress and many modifications in the actual system might take place.

Trends in electronic publishing of research in economics will be deeply influenced by many factors particularly:

- (1) People: the willingness to participate in creating and maintaining updated electronic information. One can't rely on a website to find recent publications if one is not sure of finding a stable, reliable, standardized file.
- (2) Institutions: organizations have usually sustained the publication of print working-papers. As now, the projects inventoried have always been built by a faculty member with little support from the institution. The initial eagerness of some committed faculty can't be

preserved if the institutions can't establish a standard framework to produce electronic paper series on a regular basis and according to acceptable standards

(3) Technology: certain technological features have a real potential to increase scholarly electronic publications. The software used to create electronic files are commonly adopted so one can easier download files and display their contents. The challenge is to find the valuable information within the chaos of the Internet. As noted by David Revelt "finding information is often a difficult task, and while search engines and lists of links are the most obvious attempts to mitigate this problem, no practical solution has yet been found".

(4) Libraries: information units may play an important part in order to promote the use and acceptance of electronic paper series through the enhancement of electronic issues among the research community. Intergrating references to electronic material within libraries' catalogs could be a signal to adopt electronic publication. We believe that the committment of some economists to developing new strategies in order to disseminate preprints, should rely on the library expertise to get the necessarily credit for the new product.

Back to our basic inquiry about the trends in publishing paper series in economics: The survey presented in this paper shows that producing electronic publications is, at the same time, a source of hope and eagerness for some people and the origin of skepticism and uncertainty for others. We have noted more commitment among individuals (both researchers and librarians) than among institutions, with the consequence that individuals must work hard to convince institutions about the robustness of the cyber world as an emerging arena for the dissemination of knowledge.

7) Conclusion:

The preceding chapters have exposed some issues concerning the ongoing changes with the use of electronic publishing of working-paper series in economics. While institutions' contributions to promoting and maintaining access to online information of paper series is, in many cases, limited, there is an effective interest among the economics community to built central archives in order to enhance awareness about preprints. Despite the enthusiasm about the benefits of digital publications, many scholars are still uncertain about the technological answers to academic publishing difficulties.

Beyond the actual debate, one has to notice that electronic publishing is far from a redical change in communicating research findings, but a new method to create, and disseminate information among a special community as well as among the Internet users. Electronic

facilities provided by the Internet still have problems related to finding tools; standards, establishing source authority, URL's movements and access difficulties. Nevertheless, the electronic environment "gives online users unprecedented power to follow unanticipated paths to unpremeditated payoffs",

(R. Keith Raney, 1998). Meanwhile, people will continue to enjoy both printed publications and online texts, and grey literature should take benefits from the two media.

References:

- AGORA** (<http://hosted.ukoln.ac.uk/agora/scope/htm>)
- BibEc** (<http://netec.mcc.ac.uk/BibEc.html>), Fethy Mili, Ed.
- EconWPA**, Economic Working Paper Archive, (<http://econwopa.wustl.edu>), Rober P. Parks, Coordinator
- Eisenberg, A.** (1997) "Dislinking the Internet", *Scientific American*, 6,
- Feldman, S.** (1998) "Conference Showcases the Move from Dream to Reality", *Information Today*, Vol. 15, No. 6
- Gelfand, J.** (1998) "Teaching and Exposing Grey literature: What the Information Profession Needs to Know - Examples from the Sciences." *Collection Building*, Vol. 17 No.4 19, pp. 159-166
- Goffe, W.L.** (1994) "Computer Network Resources for Economics", *Journal of Economic Perspectives*, Vol 8 No 3, pp. 97-119
- Goffe, W.L. and Parks, R.P.** (1997) "The Information Infrastructure in Economics". *Journal of Economic Perspective*, Vol. 11 No. 3, pp.75-94
- P. Grenquist**, "Why I Don't Read Electronic Journals: An Iconoclast Speaks Out," *The Journal of Electronic Publishing*, 3 (1), ISSN 1080-2711, 1997. (<http://www.press.umich.edu/jep/03-01/Iconoclast.html>)
- Karlsson, S. and T. Krichel** "RePEc and S-WoPEc: Internet access to electronic preprints in Economics. Unpublished
- Krichel, T. and Wichmann, T. (1994) "Internet Primer for Economists, *The Economic Journal*, Vol 104, (November), pp. 1496-1523
- Mackinzie Owen, J.** (1999) "Role of the Library Sector", *GreyNet Quarterly Newsletter*, Vol. 8, No.1
- NetEc** (<http://netec.mcc.ac.uk>), Thomas Krichel, Coordinator.
- Newcombe, J.** (1983) "Electronic Information Distribution. The Role of the Traditional Publisher and the Librarian", *Special Libraries*, April, pp. 150-155.
- Tunner, E., Gyeszly, S.D., and Rholes, J.M.** (1998) "Electronic and Traditional Sources for Newly Established Brach Library:: Product Availability and User Preferences". *Collection Building*, Vol. 17 No. 3, pp. 123-128
- Raney R. Keith** (1998) "Into a Glass Darkly". *The Journal of Electronic Publishing*, Vol.4, Issue 2, December 1998. (<http://www.press.umich.edu/jep/>)
- Revelt, D.** "Electronic Working Paper Standards", (<http://alfred.sims.berkeley.edu/working-paper-standards.html>)
- Shreeves, E.** (1992) "Between Visionaries and the Luddites: Collection Development and Electronic Resources in the Humanities". *Library Trends*, Vol. 40(4), pp. 579-595
- Warren-Wenk, P. and Vivienne Monty** (1996) "Internet Use Patterns and the Scholarly Research Process", in "The Challenge Face of Reference", Edited by Lynne M. Stuart and Dena Holiman Hutto. JAI Press Inc

Grey Literature Based Surveys of Information Industry Development in Central and Eastern Europe

Krystyna SIWEK

Information Processing Centre, Al. Niepodleglosci 188b, PO Box 355, PL-00-950 Warsaw, Poland

Abstract

During transformation and accession period in Central and East European countries, specially after dissemination of Bangemann Report, various surveys, reports, studies dealing with computer and telecommunication industry development appeared in form of publications, on WWW pages and as internal documents. They are prepared by external bodies as EU, OECD or global telecom organizations. Some studies have been conducted by foreign and Polish experts affiliated to universities, research units or consultancy agencies.

There is still growing number of units dedicated to information society affiliated to government bodies, regional structures of administration and NGOs. Research organizations are involved in many European projects dealing with information technologies development in CEEC. One of such projects is ESIS - European Survey of Information Society.

The Information Processing Centre as a local partner of ESIS II is responsible for monitoring and reporting the information industry development in Poland utilizing the experience in grey literature oriented information services.

INTRODUCTION

Surveys, overviews, statements, reports and numerous statistic tables and analytic documents concerning Central and Eastern Europe countries in transition period are issued as books, articles, conference proceedings, internal reports, in printed and electronic form. All these documents are a grey literature type according to redefined description during GL'97 Conference in Luxembourg.

An example of information services on political, social and business changes in CEEC is CEEBIC - Central and Eastern Europe Business Information Center of US Department of Commerce opened in 1990. It issues reports on various aspects of business in form of monthly publications and articles on CEEBIC website. Business opportunities in industries of emerging markets are analysed. Among reports concerning computer and telecommunication industry in Poland are: "IT Investment by Sector" (September 1999), "Internet Services" (July 1999), "Industry Sector Analysis of the Personal Computer Market" (March 1999), "Telecommunications Overview" (April 1999), "TPSA Privatization" (August 1999).

United Nation Organization agendas as UNESCO, UNDP, UNIDO provide support services established in special attempt to democracy and open market processes, aimed to support changes in post-communist countries, to develop their economies and industries and social

conditions as well. Needless to mention initiatives of the Open Society sponsored by Soros Foundation.

INFORMATION SOCIETY PROJECT OFFICE

After dissemination of Bangemann Report many European Information Society initiatives and projects are aimed to monitor and report information technologies development in European Union countries and in CEEC. Most important are: MISAC - Monitoring EU-CEEC Information Society Pilot Actions. GIP - G7 Global Inventory Project. TEN-34 - Trans European Research Networks. FEMIRC (A Fellow Member to Innovation Relay Centres) - a Inco-Copernicus/Innovation Programme project is one of awareness actions of Building Information Society Action Plan for CEEC. It has three action lines: information on RTD EU programmes, innovation and technology transfer, building Information Society.

The Information Society Activity Centre (ISAC) of the European Commission manages a multiannual programme to establish the Information Society in Europe. Its part is the Information Society Project Office (ISPO), which is an intermediary between EC and external relevant organizations. The other ISAC activities are Information Society Forum and international co-operation, including co-operation with Central and Eastern Europe countries. In accession countries harmonization with European Union standards and regulations is a crucial problem, so many of internal reports, statements and recommendations deal with necessity of structural changes of telecommunications. Supporting activities are based partly on PHARE programme.

Information Society Project Office develops two Information Society Inventories. One of them is mentioned above Global Inventory and the second - European Survey of Information Society (ESIS).

At the beginning of 1997, the Information Society Promotion Office of the European Commission had launched ESIS with the objective to build an inventory of projects and promotional activities. The 3rd EU/CEEC Information Society Forum that took place in Brussels in October 1997, and at the Euro-Med Net 98 Conference which took place in Cyprus in March 1998 stated to extend ESIS to Central and Eastern European and Mediterranean countries and territories. The project ESIS II started in March 1999 and will last until 2001.

METHODOLOGY OF EUROPEAN SURVEY OF INFORMATION SOCIETY

An inventory of Information Society projects - interactive projects with a societal dimension that use innovative technologies of information and communication and provide remote access. Each national contractor collects information about such projects in his country through bilingual questionnaires distributed to project promoters in printed and electronic form. The projects and action descriptions entries are sent to central database developed by management team of ESIS II. The central database is managed on the web by SQL server. An extranet web site is installed as a technical support and communication tool for management team and national ESIS partners. Every three months, the inventory of projects is updated.

Reports on regulatory developments and liberalization evolution, existing major network facilities in public utilities, statistical analysis of the IS data and indicators and overviews on key persons and organizations are prepared on national level, then separate syntheses for Mediterranean and Central and Eastern Europe countries are compiled. They are published on European Survey of Information Society Projects and Actions Web site.

ESIS II in Poland

The Information Processing Centre (OPI) is a local partner in ESIS II project for Poland. The leading partner is the Centre for Socio-Legal Studies (PCMLP) affiliated to the University of Oxford.

The Information Processing Centre, an information agency affiliated to State Committee for Scientific Research, is a publisher and databases vendor on research policy (RTD institutions, RTD projects and results etc). OPI has strong everyday cooperation with research and information units in all the universities and RTD institutes in Poland. These institutes are main information sources for OPI.

In the initial phase of ESIS II in Poland nationwide event for representatives of information technology and telecommunication institutions has been organized. Appropriate leaflets, e-mails and information on OPI server are disseminated. Training and individual consultancy is provided by OPI staff. Feedback for corrections and addenda of data between project promoters are provided as well.

There are differences in accessibility of telecommunication and information facilities. It depends on geographical region and differs from main research and industry centres as Warsaw, Cracow, Wroclaw, Poznan, Katowice and in remote regions. Specially North-Eastern part of Poland is less developed and there ESIS II survey can be conducted mostly by frequent personal visits and by standard mail.

SOURCES OF INFORMATION ON PROJECTS

Data for inventory of IS projects in Poland will be collected on Information Processing Centre server. Inventory will be conducted and developed by ESIS II in Poland team. Mostly electronic questionnaires will be used. Downloaded printed copies will be accepted as well. During two years about 100 projects and actions conducted in Poland or in co-operation with Polish partners are estimated to be included into ESIS II inventory. Estimation is based on evaluation of OPI databases on projects, experts and research institutions. Another reason of this estimation is number of projects in ESIS I database accessible on ISPO server. researchers, industry and administration. Projects under estimation started from early 1990s and were completed in 1997, 1998 or 1999. INCO-Copernicus and other European projects are the ones which started in the middle of 1990s mainly in 2 years period. KBN and university grants in average are scheduled for 2 years as well. Nation-wide projects founded by public administration or industry, depending of comprehensiveness and availability of financial support take time of 4-5 years, even longer. Increase will be stimulated by Information Society Programme in European Union Fifth Framework Programme for Research, Technology and Demonstration, in which partners from accession countries can participate on equal rules as participants from European Union, and growing interest among researchers, industry and administration.

The most frequent topics of current information society related projects identified by local ESIS II project team in the Information Processing Centre are Geographic Information Systems, CAD/CAM, technology convergence. Databases and specialised information services accessible by Internet are frequent mentioned by respondents. Projects are to be implemented in research and education, local administration, e-business, applications in various branches of industry and public services, media and entertainment, and telework. New are telemedicine, "intelligent house and office", internet services for disabled.

The estimated percentage of reports by sector is:

- 40% - Research and Education
- 20% - Media and telecommunications
- 30% - Business (including e-commerce)
- 10% - Administration

Potential respondents are identified through such sources as Information Processing Centre databases on Current and Completed Research SYNABA and POLISH RESEARCH DIRECTORY, through "INFOBAZA" inventory, Database of Grants funded by the State Committee for Scientific Research, proceedings and programmes of IS events, www services and articles in journals. CORDIS Database includes information of Polish projects conducted in framework of EU programmes.

INFORMATION SOURCES ON REPORTS

Data for reports are accessible from governmental and research documentation, Polish Press Agency, professional and scientific journals, WWW information services, specialized services as State Statistical Office, relevant databases and information services on national, regional or on European level.

Key organizations in the Information Society technologies and content sector in Poland include governmental and local administration, private business (Small and Media Enterprises), industry and services, research and technology institutions and associations, and the media sector.

Governmental information

Governmental statements on information society are those of the State Committee for European Integration, State Committee for Scientific Research, Ministry of Posts and Telecommunications and Ministry of Internal Affairs and Administration. The governmental document prepared in September 1998 by Ministry of Posts and Telecommunications is "A draft of telecommunications market development for 1998-2001" which includes the estimates of telecommunications development. Special departments or sections dedicated to Information Society development have been set up in almost all ministries. A lot of internal reports, surveys etc deal with EU integration. Special attempt is given to harmonization of the law before accession to EU, privatization of telecommunication structures, TV digitalization, electronic commerce development.

The Ministry of Internal Affairs and Administration is the national coordinator of Y2K Problem in Poland. The National Action Plan concerning Y2K. was undertaken by the Council of Ministers on 24 march of 1999 and all the actions on governmental lever are planned and monitored. The Ministry's Governmental Registers, Telecommunications and Information Technologies Department conducts the information service on Y2K problem. It includes updated reports on the government's administrative preparations for Y2K, a timetable of deadlines for Y2K preparation in power supply, telecommunications, transport, banks, finances, health and social care sectors and other related issues. A lot of "grey documents" on Y2K are disseminated as governmental documents, surveys and guidelines. The State Committee for Scientific Research (KBN) as a governmental body responsible for science in Poland prepared a document "Year 2000 - The preparation of research computer infrastructure" based on inquiries among most important WANs, MANs and LANs providers concerning software and hardware quality.

Nongovernmental organizations

Relevant associations, foundations, chambers of commerce and industry, research and professional journals usually prepare yearly reports, surveys and rankings.

The representative examples of non-governmental organizations are the Polish Chamber of Information Technology and Telecommunications, Polish Chamber of Electronics and Telecommunications, the Association of Polish Software Market "PRO", the Polish Forum of Information Society, Cities in the Internet Association, the Polish Information Processing Society, Society for Open Systems Development, Scientific Society for Information, The Polish Commercial Internet Forum.

Promotional activities in Poland

Poland is an active member of many Information Society initiatives on international, European, national, regional and local level. The State Committee for Scientific Research (KBN) coordinates the growth of the Polish Internet and supports development of computer and Internet facilities for scientific community in Poland. On KBN home page, National Contact Point for Global Information Society service is conducted by Information Technology Systems department.

Main conferences and exhibitions, both international and nationwide, dedicated to computer industry are "Computer-EXPO", "COMNET", and "POLMAN", "Infosystem" conference, "Komtel" and "INFOBAZY'99".

Publications

Following the First Congress of the Polish ICT in December 1994, the Polish report was compiled and focused on the most important issues of IT development in Poland. The Second Congress, 30 - 2 December 1998 in "Treaty for building the Information Society in Poland" stresses the challenge for all the citizens. "Information Technology Development in Poland" report after The Second Congress is available in a printed version and on the Internet underlines obstacles of information society building in Poland.

Among computer and telecommunications magazines and journals in Poland are "Teleinfo", "Computerworld", "Internet", "Enter", "Network". Many of them provides yearly rankings of companies. "Information Society Forum" ("Forum Społeczeństwa Informacyjnego") is an appendix to "Teleinfo".

The biggest computer and telecommunication companies active on the Polish market are mainly branches of giant foreign companies but the percentage of Polish businesses is still growing. In the new edition of "TELEINFO 500 - Polski Rynek Teleinformatyczny 1998. Raport", edited in June 1999, the 600 largest Polish computer companies, 200 computer companies operating in Poland, 150 largest Polish telecom companies and 25 top telecom companies operating in Poland in 1998 are listed.

Electronic sources

The expansion of web services in Poland is significant. The most popular to date are basic services such as e-mail and WWW pages - which are mainly targeted at institutional promotion. More advanced services such as electronic commerce are at the primary stage. The longest established are electronic book stores as "Ksi" Gki OnLine" (Books OnLine) - a part of Polska Online portal. A number of portals is still growing. The next is Onet.pl of Optimus S.A. and new portals of Internet Venture Poland (IVP) and Telekomunikacja Polska S.A.

For reports on regulations web pages of the Polish Parliament, The Ministry of Posts and Telecommunications, Ministry of Internal Affairs and Administration, The State Committee for Scientific Research and Committee for European Integration are frequently browsed. Every telecom operator and Internet provider has established web service where up to date information on new technologies, telecommunication services, tariffs etc is presented. Databases are provided by research

institutions and libraries via internet. Information Processing Centre conducts on its own internet server bilingual databases as SYNABA, Dissertations and Polish Research Directory, accessible free of charge for remote users.

Conclusions

ESIS II survey can be a virtual grey literature collection for a growing number of noncommercial publications, internal documents and electronic multimedia resources concerning information and telecommunication market in Central and East Europe. Together with such a collection of documents and records gathered from EU countries during the first phase of ESIS project and with Mediterranean countries information resources it seems to be a very comprehensive inventory of European contribution to worldwide challenge in computer and telecommunication technologies. Much more effort should be given to disseminate these valuable knowledge resources broadly and provide access in "user friendly" way for different user groups of the global information society.

REFERENCES

Third EU/CEEC Information Society Forum. Brussels, 9-10 October 1997. Proceedings, European Parliament/European Commission, Brussels

Building the European Information Society for us all. Final policy report of the high-level expert group. (1997). European Commission. Brussels.

Coopers & Lybrand Technology Consultants (1998) *Impact of the enlargement of the European Union towards the associated central and eastern European countries on RTD-innovation and structural policies. Final report.* European Commission, Brussels

INFOBAZY'99. Bazy Danych dla Nauki. Gdansk [30 August - 1 Sept. 1999, Proceedings] Politechnika Gdanska, Gdansk

Polish Research Directory 1999. Information Processing Centre. Warsaw. Available http://www.opi.org.pl/en_baz.htm

Scientific Research in Poland. (1999) Information Processing Centre. Warsaw. Available http://www.opi.org.pl/en_baz/htm

Market Research. Poland. Central and Eastern Europe Business Information Centre. US Department of Commerce. Available <http://www.mac.doc.gov/eebic/country/poland/market2.htm>

Ministerstwo Łacznosci (1998) *Zarys strategii rozwoju rynku telekomunikacyjnego na lata 1998-2001*". Available <http://www.ml.gov.pl>

"Computerworld Online" Available <http://www.computerworld.pl>

"TeleInfo OnLine" Available <http://www.teleinfo.com.pl>

Use Pattern of Grey Literature in Rural Development An Indian case study

By T. Rama Devi & Anil Takalkar

Centre on Rural Documentation, National Institute of Rural Development
Rajendranagar, Hyderabad – 500 030 India

ABSTRACT

Rural Development Literature in India is growing as the expenditures meant for the sector are increasing under every Five Year Plan. There are number of institutions, NGOs and individuals undertaking research on various aspects of development. National Institute of Rural Development, Hyderabad, India is the apex research, training and consultancy organisation in India on rural development. It publishes several periodicals, monographs, proceedings of workshops and seminars, case studies etc. Journal of Rural Development (JRD) published by NIRD since 1982 covers subjects like agriculture, economics, political science, psychology, sociology, management and gender studies etc. This refereed and peer reviewed journal carries both theoretical and empirical research based articles pertaining to India and other developing countries. Besides, it publishes case studies, research notes, book reviews etc.

The present study examines the distribution, classification, and categorisation of articles published in JRD during 1982-1997 and the number of citations relating to grey literature. Based on the method of citation analysis, the use pattern, accessibility and availability of grey literature to the contributors of the articles were studied. Also an analysis of the formats of the grey literature pertaining to India covered in the two International Abstracting Services, Rural Development Abstracts by CABI, U.K. and Abstracts of Tropical Agriculture by Royal Tropical Institute, Netherlands was made and compared with that of JRD for the period 1982-97.

Study concludes that citation pattern is similar in theoretical and empirical articles. Increasing number of citations of grey references indicates the importance of and demand for grey literature. In order to help improve their access and awareness and consequent supply to enhance their utility, an action plan is suggested. Under this plan, a special programme of networking of socio-economic institutions in India especially the 25 State Institutes of Rural Development (SIRDs), spread across the vast country is to be initiated.

As a part of this activity, various government agencies, research and training institutions, NGOs, universities, banking institutions are being requested to supply their publications / list of publications which will form an input to the network. A separate database of Grey literature is being created at NIRD to make it available through Internet (www.nird.org).

INTRODUCTION

Grey Literature (G.L) plays an important role, as a means of information dissemination, needs no emphasis. The importance of such fugitive documents was brought to international attention by the report of the DEVSIS Study team (IDRC, 1975). To quote the report, " the G.L is made up of unpublished working papers, feasibility and pre-investment studies, theses, research reports and documents of governments and international organisations that are not widely

disseminated". The main characteristics attributed to such fugitive or grey literature are non-availability through normal trade channels, limited readership and limited distribution.

Development in general and rural development in particular assumes prime position on the agenda of activities of governments of various developing countries including India. Various inter-governmental organisations, non-governmental organisations, research and training institutions, development aid agencies etc. all over the world are engaged in rural development activities. Development initiatives of these organisations are manifest in a number of strategies and programmes. In the process, these organisations generate enormous literature that is mostly in the form of grey literature. According to an estimate, 5000 documents per year are produced in India alone on aspects directly relating to rural development by various institutions and organisations (Raju, K.A. 1994).

In view of the increasing output of G.L in the field of rural development, it is often assumed that it is also being widely gathered, collated, documented or processed for use by the development community engaged in planning, policy making, training or implementing rural development programmes and projects.

As stated by the DEVSIS report major users of any development information system are those people who are themselves involved in the mission. It has been observed that the generators of literature in the development science are themselves the users of the system. Interestingly, this phenomenon is more conducive for reprocessing, repackaging for effective dissemination in a networked environment.

In India (as also in many developing countries), "development" is achieved through the planning processes which are commonly known as Five-Year Plans. The successive Five-Year

Plans initiated in 1951-52 in India are the major instruments for achieving the goals of development. Currently Rs.100 million are earmarked every year towards rural development which is slated to be doubled during the ongoing Five Year Plan (1997-2002). In formulating development plans, there is the need for a wide range of information covering many disciplines. Simultaneously, many institutions came into existence to undertake development research to provide new insights into various problems like growing poverty, unemployment, food security and for taking appropriate decisions. Generation of new knowledge through research in development is an essential pre-requisite and forms the basis of formulating and implementing development plans and programmes. National Institute of Rural Development at Hyderabad is one such leading institution that came into being in the beginning of 60's to undertake development research and training and help provide decision support to the development community.

National Institute of Rural Development (NIRD) is an apex organisation under the Ministry of Rural Development, Government of India. The mission of NIRD is to facilitate rural development in India to improve the quality of life of the rural poor through various programmes and projects. In order to achieve this mission, NIRD undertakes training, research and consultancy and information dissemination activities.

As a result of its varied activities over the years, NIRD has developed a vast reservoir of contemporary knowledge in rural development. As a part of its information dissemination function, NIRD brings out Journal of Rural Development, Newsletters, Research Highlights, Recommendations of Seminars and Workshops, Research Reports, Monographs, Occasional papers, etc.

JOURNAL OF RURAL DEVELOPMENT (JRD)

Journal of Rural Development is an official journal of NIRD and is seen as a representative journal in the field of rural development in India. NIRD was established in the beginning of 60's under the name of National Institute of Community Development (NICD) to reflect the then government's policy and activities on community development. As a vehicle to disseminate information about its contributions to the field, two journals were simultaneously being published. i) Behavioural Science and Community Development and ii) Panchayati Raj Digest. In the middle of 70's under the thrust of second development decade, a new policy paper on Rural Development was released by World Bank in the year 1975. Integrated rural development as a policy was initiated and India also adopted the same. In keeping with the developments, the name of NICD was changed to National Institute of Rural Development (NIRD) by the end of 70's. Consequently NIRD merged the two journals into one and started issuing the JRD from the year 1982 to 1997. Thus JRD in a way reflects the transition of development policies pursued by India. It acquired slowly a prestige and standing not enjoyed by any other contemporary Indian journal in the complex subject of rural development.

Rural development being interdisciplinary in nature, the coverage of the journal includes subjects like agriculture, economics, political science, local government (Panchayati Raj), psychology, sociology, management, gender, environment, energy, poverty, and industry. This refereed and peer reviewed journal carries both theoretical and empirical research based articles pertaining to India and other developing countries. Besides, it publishes case studies, research notes, book reviews etc. The journal aims at promoting study and research in rural development. It seeks to uncover links between the social sciences and rural development and to forge and strengthen them where necessary. It provides a forum for exchange of views between various social science disciplines and the policy makers, planners and the executives concerned with rural

development. Hence, for the purpose of the study, we have selected JRD since its inception 1982-97.

OBJECTIVES:

The objectives of the present study are

1. To examine the use pattern of grey literature with special reference to Rural Development.
2. To compare the coverage of grey literature on India in international abstracting services
3. To suggest an action plan for effective bibliographic control of grey literature.

METHODOLOGY

As the aim of our study is to understand the pattern of usage of grey literature by the social scientists with special reference to the faculty of NIRD, we have examined the data relating to the type of articles that are published in the journal during 1982 to 1997, their number, their focus (empirical or otherwise), the number of citations, the category of citations and their distribution together with many other details like classification of the subjects of the given citations, categorisation of the different author groups and the contribution of NIRD faculty.

As the citation analysis is one of the accepted methods to assess or determine the actual use of documents and resources, this method was chosen to investigate the use of grey literature. Analysis of such citation in a given scientific communication can reveal useful information like the relative use of different categories of documents such as conferences/ seminars / workshop proceedings, research reports, Committee and Commission reports, case studies, annual reports, dissertations etc. Additionally the data that was derived from JRD pertaining to grey literature was compared with that of the two international abstracting services 'Rural Development Abstracts' of

CABI, U.K. and 'Abstracts of Tropical Agriculture (TROPAG)' brought out by Royal Tropical Institute, Netherlands. We have also seen from this two international abstracting journals, the coverage of grey literature for the period 1982-97 pertaining to India. This was done in order to know the extent of coverage as well as the categorisation of materials to compare with the data obtained from JRD.

DATA ANALYSIS

JRD was published as a bimonthly during 1982-93 (12 years) and from 1994, the periodicity was changed to quarterly. Thus a total number of 88 (72+16) issues were checked and data collected. It is observed that a total of 596 articles were published during the 16 years. From these (a) articles based on empirical studies (b) articles contributed by single author and belong to NIRD or otherwise (c) articles with citations were differentiated. The following table gives the distribution as stated above :

Table I: Distribution of articles according to the given category

	Total No. of Articles : 596	
	Number	Percentage
Empirical based Studies	283	47.4%
Non-empirical based studies	313	52.6%
With citations	486	81.6%
Without citations	110	18.4%
Contributed by single author	333	55.8%
Contributed by more than one author	263	44.1%
Articles by NIRD Faculty	241	40.43%
Articles by Non-NIRD	355	59.5%

The above data gives an idea that there is more or less an equal distribution of the empirical and non-empirical based articles. The articles contributed by single author and those coming from outside NIRD are significant. We can infer that the journal's standing among the development community is increasing over the years attracting contributions from outside NIRD and most of the articles are having citations, that include a sizable grey literature.

CATEGORIES OF CONTRIBUTORS

We have examined the categories of persons who contributed the articles from NIRD and outside NIRD. The total number of contributors from NIRD is 362 and the rest are from different organisations. The following table gives the categorisation of the authors of the articles from NIRD and others. The number of contributors do not tally with the number of articles as some of the articles are written by more than one author.

Table II: NIRD and Non-NIRD Contributors

NIRD Contributors

Senior Professionals	Middle level	Junior Faculty	Total
151 (41.7%)	126 (34.8%)	85 (23.4%)	362

Non-NIRD Contributors

Academicians	Functionaries	Planners	Administrators	Others	Total
518 (86.9%)	48 (8%)	41(6.8%)	(2.68%)	7 (1.01%)	630

The category of senior professionals from NIRD consists of Directors, Deputy Directors of different centres. Faculty members like Assistant Directors, Senior Research Officers are categorised as middle level professionals and Research Associates are grouped into junior faculty.

As regards non-NIRD contributors, those from research and training institutions, and universities are grouped as academicians. Government functionaries like Block Development Officers, Project Directors of District Rural Development Agencies (DRDAs), are categorised as functionaries, officials associated with planning and designing of development programmes, Chief Executives of different institutions have been categorised under planners. IAS officers who have contributed in a significant number have been categorised separately as administrators. Contributors who have not mentioned their designation or affiliation and could not be classified into any of the given groups, have been categorised as others.

The occupational category-wise distribution of the NIRD contributors shows that the senior faculty is contributing more often to JRD than the other categories. This can be attributed to the fact that senior faculty in NIRD are involved relatively more in research activities than the juniors. Also the projects that are designed in NIRD are directly related to either the rural development programmes of the Ministry of Rural Development, Government of India or of the thrust areas of the institution. In both the categories, the senior faculty are more involved. As a result, the outcome of the projects and the programmes heavily depended upon them.

As regards the distribution of outside contributors is concerned, the academic community occupy the first position followed by the others. Most of these academicians belong to the universities and other development oriented research and training institutions of the country. As stated earlier, the enormous development funds that are spent in the name of rural development by the Central and State governments increased the number of projects undertaken resulting in the production of different project reports and other attendant publications. These are mostly handled by the academicians which is reflected in their contribution. It is also seen from the data that authors from different backgrounds and institutional affiliations contributed to the output of literature in varying degrees.

Table III : Distribution of citations and percentage of grey references

	1982-86	1987-91	1992-97	Total
Total no. of cited references	572	2292	1903	4767
Total no. of grey references among the cited	148	597	581	1326
Percentage	25.8%	26%	30.5%	27.8%

The Table-3 shows distribution of articles in a five year interval along with the number of citations and the percentage of grey references among the cited. The Table-4 gives the idea about the grey references used by NIRD faculty. It is interesting to know that there is an increasing trend in the citation of grey literature progressively over the years. This can be attributed to two factors : the increasing production of grey literature and its relative availability. Similarly, the citations of grey literature by the faculty of NIRD reveal that the senior faculty perhaps has more awareness as well as access and availability of the grey literature than the others. Senior Faculty, by virtue of having information about the various programmes and projects that are ongoing in the country in their field of specialisation as well as due to their frequent visits to the headquarters of the several ministries in the state and central government, are relatively well off in their awareness of different kinds of material. Also 'invisible college' concept works equally well with Social Scientists.

Table IV : Type of Grey Literature used by NIRD Faculty

	Senior Professionals (Dir.,D.D.- 262)	Middle level Faculty (SRA, A.D-160)	Junior Professionals (R.A's-45)
Conferences/ Workshop	50 (19%)	19 (11.8%)	6 (13.3%)
Committee / Commission Reports	57 (21.7%)	24 (15%)	8 (17.7%)
Case Studies	7 (2.6%)	2 (1.3%)	1 (2.2%)
NIRD Reports	29 (11%)	30 (18.7%)	5 (11%)
Dissertations	18 (6.8%)	1 (0.6%)	2 (4.4%)
Others *	101 (38.5%)	84 (52.5%)	23 (51.1%)
Total	262	160	45

* Annual Reports, Project Reports, Survey reports, Review papers, Evaluation reports, Progress report, guidelines etc.

SUBJECT COVERAGE OF GREY LITERATURE

We have also analysed the kind of subjects the cited grey literature cover as well as their type. The following table gives the various subjects covered in the articles contributed to JRD as well as the corresponding percentage of citation belonging to that topic of the contribution.

Table V : Subject Coverage of Grey Literature

Subject of Articles		Subject of Citations given in the articles	
Rural Development	116 (19.4%)	Rural Development	231 (17.4%)
Agriculture	90 (15%)	Agriculture	135 (10.1%)
Local Government/ Panchayati Raj	65 (10.9%)	Local Government	147 (11%)
Gender	38 (6.3%)	Gender	55 (4.1%)
Industry	31 (5.2%)	Industry	40 (3.0%)
Forestry	4 (0.6%)	Forestry	43 (3.2%)
Poverty	30 (5.0%)	Poverty	40 (3.0%)
Total	596		1326

Note : Remaining subjects like energy, employment, livestock, watershed, finances and many others covered are not given here for want of space.

JRD by virtue of its stated focus attracted articles more in number that relate to rural development followed by allied subjects like agriculture, local government, and gender studies. Citations also follow more or less the same pattern.

COMPARISON WITH OTHER SERVICES

In order to understand the kind of accessibility and availability of grey literature for the purposes of citation as well as bibliographic control, we examined two major international abstracting services namely Rural Development Abstracts and Abstracts of Tropical Agriculture. We have seen the coverage for the same period 1982-97 by using the CDs available under the name AgECONCD and TROPAGCD respectively. We have also restricted our search from the

CDs to the grey literature on India without any restrictions on place of publication. The following table gives the distribution of grey citations as against the total references in case of JRD and total entries in case of two CDs pertaining to India.

Table VI : Coverage of grey literature in different databases.

	JRD	AgECONCD	TROPAG
Total references for the period 1982-1997	4767	21887	14329
Grey references	1326 (28%)	3803 (17.3%)	1241 (8.6%)

It can be seen from the above table that the coverage of grey literature is insignificant, bringing home the point that there is a lack of bibliographical control of G.L. The findings are supported by a comparative study of three international abstracting services viz. Rural Development Abstracts, International Development Abstracts and Abstracts of Rural Development in the Tropics conducted earlier. The study showed that the coverage of grey literature in these abstracting services was inadequate compared to the other types of literature abstracted. It was also observed that the extent of duplication among the services was very insignificant (Raju, K.A. 1991)

Table VII : Types of materials covered under grey

	JRD	AgECON	TROPAG
Conference Proceedings	208 (15.6%)	535 (14%)	406 (32.7%)
Research Reports	44 (3.3%)	37 (0.9%)	3 (0.24%)
Thesis / Dissertations	102 (7.6%)	198 (5.2%)	10 (0.8%)
Case Studies	45 (3.4%)	872 (22.9%)	326 (26.2%)
Annual Reports	38 (2.8%)	671 (17.6%)	23 (1.8%)
Guidelines	10 (0.75%)	160 (4.2%)	51 (4.1%)
Manuals	16 (1.2%)	330 (8.6%)	104 (8.3%)
Project Reports	31 (2.3%)	5 (0.13%)	2 (0.16%)
Training programmes	6 (0.45%)	90 (2.3%)	36 (2.9%)
Others *	826 (62.2%)	905 (23.7%)	280 (22.5%)
Total	1326	3803	1241

* Working papers, Action plans, Evaluation Reports, Issue papers, Occasional papers, State of the art, Translation progress reports etc.

We have also seen the different types of grey literature (Table 7) such as conferences / seminars / workshop reports, research reports, case studies, thesis, dissertations, annual reports etc. covered in order to know the reach of these abstracting services in the bibliographical control of the grey literature emanating from India.

It is observed that the abstracting services are able to cover more the conferences/ seminars literature, case studies, dissertations. This may be because that the conferences and seminars are publicised more by the organisers nationally and internationally and the primary periodicals cover more case studies, list of dissertations. Also there are several other directories available which give the list of various organisations that operate in rural development. Information about the organisations facilitates getting annual reports relatively easier than any other documents. As a result, we find more coverage of such documents in the CDs and also their accessibility is reflected by their citation in the contributions to the JRD. However in the final analysis, there is still a need to have an indepth study of other abstracting services of national and international organisations in order to study the coverage of the grey literature, and their use pattern. Meanwhile, we are implementing an 'action plan' to increase the awareness and accessibility of grey literature taking advantage of the advances in information and communication technologies.

ACTION PLAN

1. A Computerised Library and Information Clearinghouse (CLIC) has been setup at CORD with a view to provide information on different aspects of rural development at a centralised location. The documents / information collected under CLIC falls mostly in the category of grey literature such as research reports, evaluation reports, case studies, success stories, conference / seminar proceedings, etc. Special efforts would be made to enrich the grey literature collection by approaching various government agencies, research and training institutions, NGOs, Universities, banking institutions and the like, who are the generators of Grey Literature. CLIC is presently accessible on internet (URLs: www.PanAsia.org.sg/nird and www.nird.org)

2. State Institutes of Rural Development (SIRDs) are the apex training and research institutions at state level. There are 25 such SIRDs. These are being brought into a network (Annexure-I) and would be assigned the task to collect the rural development literature generated at the state level and also in regional languages. The SIRDs would act as the depositories for such literature and send the information to CLIC for wider dissemination. The CLIC is programmed to work as a focal point for the grey literature on rural development at the national level on the lines of the European SIGLE.

3. Internet would be searched for web sites relevant for the development activities. Material of interest from these sites, especially the grey literature would be identified and downloaded and made available under CLIC for the benefit of the development community.

4. International abstracting services in the discipline, which cover grey literature would be monitored on a continuous basis to identify the grey literature covered and to procure the same as far as possible.

CONCLUSIONS

It can be concluded from the above analysis that grey literature in various types, even in the present day context continues to be one of the important media of transfer of information and a source of recorded knowledge but eludes comprehensive bibliographical control. Increasing number of citations of grey references indicates the importance of and demand for grey literature. Though theoretical and empirical studies form distinct types of intellectual contributions, they follow similar pattern of citation of grey literature according to our study. It is observed that only those publications are usually cited most which are published in the preceding five years. The awareness about and accessibility to grey literature is responsible to a considerable extent as can be noted from the citations. At the national level, inadequate bibliographical control of the grey literature is one of the major factors for this situation. Whereas at the global level as evidenced from the databases like RD Abstracts and Abstracts on Tropical Agriculture, the coverage of grey literature though appears to be numerically more but compared to other references in the databases, it is not very significant.

The implication of the study is that there is an urgent need at the national level, to coordinate the efforts of all major institutions/individuals in a given discipline which contribute significantly in the generation of grey literature. A centralized identified agency may be assigned with the responsibility of coordinating these activities on the lines of the European example SIGLE. This to a considerable extent may establish bibliographical control over grey literature and improve the awareness about, access to and consequent supply of grey literature among the concerned at the regional and national level. The initiatives of different countries in this regard may, in turn form the inputs to the agencies which are engaged in development of databases in respective disciplines with a world wide coverage to achieve bibliographic control at the global level.

With the advent of Internet, the concept of development as a global concern acquired new meaning. Many institutions are developing their own websites, homepages in order to publicise their activities, expertise, publications and products. These will definitely improve the awareness and accessibility of grey literature. Simultaneously several networks at regional or local level are emerging on the Internet, specialising in making available information generated by the participating institutions of different countries. One such is Pan Asia Network (PAN) sponsored by IDRC, Canada. Through PAN, the International Development Research Centre is aiming at promoting collaboration in research and development by connecting individuals and institutions for knowledge sharing across Asia. NIRD is also a member of PAN and it is deriving the benefit from the co-members of network in order to access information on grey literature and other useful network groups.

ACKNOWLEDGEMENT

We acknowledge with thanks Mr. R.C. Choudhury, Director General, NIRD for giving permission to submit the paper to the GreyNet 99 Conference. We also wish to acknowledge Mr. L.S. Ramaiah, Ex – Librarian of Central Institute of English and Foreign Languages, Prof.N.Laxamana Rao, Dept. of LIS, Osmania University, Hyderabad and Mr. K. A. Raju, Director (CORD), NIRD for their valuable comments and suggestions on the paper before submission to conference. The help rendered by Mr.P. Sudhakar in the tabulation of vast data before its analysis is gratefully acknowledged.

REFERENCES :

1. DEVSIS Study Team . "DEVSIS- Preliminary Design of an International Information System for the Development Sciences ". Ottawa, IDRC, p.18, 1976.
2. Raju, K.A. : Rural Development Literature : Preliminary results of a comparative study of three global abstracting services. IAALD Quarterly Bulletin, 36(4), 1991, 227-231p.
3. Raju, K.A. : Rural Development Literature : A Quick Estimate, Paper presented at the Review Meeting of an IDRC project on DEVINSA, New Delhi, 1994.
4. Commission of the European Communities, Brussels. "An up-to-date System for Information on Grey Literature produced in Europe (SIGLE)". A brochure, n.d.

ANNEXURE – 1

Figure 1 : Proposed Network of State Institutions of Rural Development (SIRDs) Connected



APARD, Hyderabad, (Andhra Pradesh)	SIRD, Shillong (Meghalaya)
SIRD, Itanagar (Arunachal Pradesh)	SIRD, Aizawat (Mizoram)
SIRD, Guwahati (Assam)	SIRD, Kohima (Nagaland)
BIRD, Patna (Bihar)	SIRD, Bhubaneswar (Orissa)
SIRD, Ahmedabad (Gujarat)	PUN-SIRD, Nabha (Punjab)
HIRD, Nilokheri (Haryana)	HCM-RIPA, Jaipir (Rajasthan)
HIPA, Shimla (Himachal Pradesh)	SIRD, Karfector (Sikkim)
IMPARD –Srinagar (Jammu & Kashmir)	SIRD, Maraimallainagar (Tamil Nadu)
ANS-SIRD, Mysore (Karnataka)	SIRD, Agartala (Tripura)
SIRD, Kottarakara (Kerala)	DDU -SIRD, Lucknow (Uttar Pradesh)
MG-SIRD, Jabalpur (Madhya Pradesh)	SIPRD, Calcutta (West Bengal)
YASHADA, Pune (Maharashtra)	SIRD, Goa
SIRD, Imphal (Manipur)	

How do various Fugitive Literature Searching Methods Impact the Comprehensiveness of Literature Uncovered for Systematic Review?

Diane Helmer BFA MLIS
Isabelle Savoie MD MHA
Carolyn Green MSc

BC OFFICE OF HEALTH TECHNOLOGY ASSESSMENT

INTRODUCTION

The systematic review is a vital tool for health care providers, researchers, and policy makers, allowing efficient and effective review of research evidence on which appropriate policy or treatment decisions can be based. The systematic review uses explicit methods to evaluate and synthesize information in a manageable form so the current state of knowledge on the effectiveness of an intervention can be understood (Chalmer, 1995).

For researchers conducting systematic reviews and aiming to reduce publication bias, the nature of fugitive literature presents a number of problems not the least of which is to define its limits. Although considerably debated, the most authoritative definition of fugitive literature remains broad (GL'97, 1997). In addition, with the advent of advanced information technologies and the Internet, new forms of fugitive literature are being introduced and becoming more widely used (Caroll, 1997). As well, more organizations and writers have access to desktop publishing software. Overall, the fugitive literature is increasing at an exponential rate. In fact, according to one estimate, its rate of growth is three to four times that of conventional literature (Farace, 1997).

Some countries, particularly the US and the UK, have developed their fugitive literature resources and made them easier to access. This is not the case in many other parts of the world. Where no centralized agency exists to assist in its distribution, the acquisition of grey literature remains problematic (Alberani et al, 1990).

Although the accessibility of various kinds of fugitive literature and document delivery has been facilitated by the development of the Internet and resource-sharing initiatives between libraries, increased access to information through these means has only partly kept pace with the mass of information.

To establish standards for searching the literature a number of guidelines have been produced (Mulrow et al, 1997)(Goodman C, 1993).

Search protocols based on these guidelines will generally require researchers to:

- search the electronic literature databases;
- search the grey literature;
- hand-search key journals;
- scan reference lists;
- identify conference proceedings;
- consult other researchers in the topic area.

The assumption has been that this series of steps will reduce publication bias and improve the ultimate credibility and precision of the systematic review (NHS Centre for Reviews and Dissemination, 1996)(Stern et al, 1997). To date however, very little evidence appears to have been gathered to show the effectiveness of some of these steps.

In the light of these difficulties, this study was designed to determine if and to what extent various fugitive literature searching methods affect the comprehensiveness of literature uncovered for systematic reviews.

METHODS

The British Columbia Office of Health Technology Assessment (BCOHTA) at the University of British Columbia undertakes health technology assessment research, whose primary purpose is to examine scientific evidence on the effectiveness of current and developing health technologies. BCOHTA bases its appraisals on comprehensive evaluations of the scientific evidence, and consequently, the literature searches for each project can be intense.

In 1993, BCOHTA adopted the use of databases to manage the bibliographies associated with each project. Since 1993, the Office uses Dbtext Works by Inmagic Software Inc. to produce bibliographies.

In undertaking systematic review, a BCOHTA search is normally divided in two parts: the electronic literature (or Level I search) and the fugitive searches (New Zealand Health Technology Assessment, 1999).

The electronic literature search comprises detailed search strategies in databases such as Medline, Embase, HealthStar, Biosis, PsychLit or Current Contents. The selection of the databases is directly dependent on the topic of the systematic review. When the electronic literature search is complete, BCOHTA researchers select articles that meet the pre-defined inclusion and exclusion criteria, and this material is retrieved.

The second part of BCOHTA's search strategy is the fugitive search. In the fugitive search:

- subject specific and specialized databases are identified and searched;
- the Internet is searched by using meta-indexes such as OMNI. [2] Search terms similar to those in the electronic search strategies are applied to these meta-indexes. A select number of search engines are used to locate organizations and researchers;
- relevant non-indexed and indexed journals are identified. These journals are then handsearched to ensure comprehensiveness.
- The search for fugitive literature also makes use of directories to indicate organizations undertaking topic-specific research. From this point, relevant organizations and researchers can be identified and contacted.

For the purpose of this study, two systematic review projects were used as sources of data. The projects chosen were topics for which it was feasible both to apply a prospective methodology and to undertake a substantive fugitive search.

The systematic reviews aimed to gather as many randomized controlled trials as possible on their respective topics. A prospective analysis was applied to the literature items retrieved by both electronic searches and fugitive searches.

The fugitive search strategies for each project are detailed in **Tables 1 & 2**. For the electronic search, a strategy designed to be consistent across both projects was applied to Medline, Embase and Current Contents.

Items meeting the inclusion criteria for each project were retrieved and entered in a BCOHTA project-specific database using DbText. Each item was coded to describe how it was found. Items previously uncovered through the electronic literature search were downloaded directly from their respective database into DbText, and the unique identifiers assigned by Medline and

similar databases allowed such records to be differentiated from the fugitive records. A second code was assigned to describe the type of item uncovered. The item types were coded independently by one librarian and by one researcher. The coding results were then compared and any differences between the researcher and librarian's coding were resolved by discussion.

PRELIMINARY RESULTS AND DISCUSSION

In total, 1034 items were retrieved. Of these, 302 (29.2%) were identified through the fugitive search. Given the difficulties encountered in acquiring the fugitive literature, this may still represent an underestimation of the effectiveness of fugitive searches. For example, not all researchers have responded to our inquiry regarding trials and other relevant information.

Most of the fugitive literature uncovered takes the form of traditional reviews. The indexing and precision of randomized trials in the specialized databases may explain this finding. In addition, fugitive items are less likely to offer abstracts. Therefore, it is often harder to determine if the items meet the inclusion and exclusion criteria before ordering the item. The search of specialized databases was the most effective method of identification, followed by the review of reference lists.

As stated above, these results are preliminary. We are still in the process of following-up on a few letters that were sent. Also, some of the items selected have been requested through inter-library loan have not been received yet. We are waiting for a response from the Inter-Library Loans department to confirm whether or not these items are available.

CONCLUSIONS

The preliminary results show that the fugitive search can be an effective tool for uncovering material for the systematic review. The quality of the fugitive literature items uncovered as well as the likelihood of that literature affecting the results of the systematic review, is yet to be determined.

REFERENCES

- Alberani V, de Castro Pietrangeli P, Rossi Mazza AM (1990) "The use of grey literature in health sciences: a preliminary survey", *Bulletin of the Medical Librarians Association*, Vol 78 No 4, p.358
- Caroll BC, Cotter GA (1997) "A new generation of grey literature: the impact of advanced information technologies", *Publishing Research Quarterly*, Vol 13 No 2, pp. 5-14
- Chalmer I, Altman DG, editors (1995) "Systematic reviews" BMJ Publishing Group, London, pp.1-2
- Farace DJ (1997) "Rise of the phoenix: a review of new forms and exploitations of grey literature" *Publishing Research Quarterly*, Vol 13 No 2, pp. 69-76
- GL'97 Luxembourg Convention (1997) "Perspectives on the Design and Transfer of Scientific and Technical Information" Third international conference on grey literature. Luxembourg, 13-14 November 1997
- Goodman C (1993) "Literature searching and evidence interpretation for assessing health care practices" Swedish Council on Technology Assessment in Health Care, Stockholm, pp. 16-32
- Mulrow CD, Oxman AD, editors (1997) "Cochrane Collaboration handbook" In: The Cochrane Library [database on disk and CDROM]. Oxford, Update Software, 1997, Issue 4
- New Zealand Health Technology Assessment (1999). "NZHTA search protocol" Available from: URL: <http://nzhta.chmeds.ac.nz/nzhtainfo/protocol.htm>
- NHS Centre for Reviews and Dissemination (1996) "Undertaking systematic review of research on effectiveness : CRD guidelines for those carrying out or commissioning reviews. CRD report; no. 4. The Centre, York, pp. 2-3
- Stern J, Simes RJ (1997) "Publication bias: evidence of delayed publication in a cohort study of clinical research projects" *British Medical Journal*, No 315, pp. 640-5

ENDNOTES

- [1] <<http://chid.nih.gov/subfile/subfile.htm>>
- [2] <<http://www.omni.ac.uk>>

TABLE 1. Fugitive Search – Lipid Lowering Drugs

Commercial databases	Web library catalogs (using Library of Congress or National Library of Medicine (MeSH) subject headings)	Internet peer-reviewed sites	Internet search engines
1. Cochrane Library 2. HSTAT (technology assessment guidelines) 3. HSRProj (MLM) 4. Dissertation Abstracts 5. Article 1st (OCLC) 6. Papers 1st (OCLC) – conferences and paper abstracts 7. TRIP database (evidence-based medicine) 8. NTIS Database 9. CRISP (Computer Retrieval of Information on Scientific Projects) 10. HTA Database 11. LILACS	1. UBC Library Catalog 2. BC Ministry of Health Library Catalog 3. Canadian Institute of Scientific and Technical Information (CISTI) catalog 4. Belinda Database (Buckinghamshire Health Authority Library) 5. HealthPromis (UK web catalog of health promotion) 6. National Health Information Center – Health Information Resource Database 7. CHID Online (Combined Health Information Database) 8. WorldCat 9. GAO Web Catalog 10. COPAC (Union catalogs in the UK) 11. NLM Locator Plus	1. UK Academic Web Directory 2. UK Social Science Information Gateway 3. OMNI (Organising Medical Networked Information) 4. Medical Matrix 5. Health Communications Network 6. Global Health 7. Health Index 8. Medweb Public Health 9. Medscape	1. Northern Lights 2. Altavista 3. Excite
In-house databases	Directories	Organizations contacted	Journals hand-searched
1. In-House Catalog 2. Analytics Database	1. ECRI HealthCare Standards 2. UHC Technology Assessment Monitor	1. US National Institutes of Health 2. University of Ottawa Heart Institute 3. Montreal Heart Institute 4. National Heart Lung and Blood Institute 5. Health Heart Program St Paul's Hospital, University of British Columbia 6. World Health Organization 7. International Task Force on Coronary Heart Disease 8. International Society of Atherosclerosis 9. Institute of Clinical Evaluation Sciences 10. American Heart Association 11. National Cholesterol Education Program 12. College of Pharmacy and Faculty of Medicine, Dalhousie University 13. American College of Cardiology 14. Ontario Ministry of Health 15. Conseil d'Evaluation des Technologies du Santé 16. Australian Institute of Health and Welfare 17. University Hospital Consortium 18. Health Services Utilization Research Commission 19. Trent Institute for Health Services Research 20. Canadian Coordinating Office of Health Technology Assessment 21. International Society for Pharmacoeconomics and Outcomes Research 22. Therapeutics Initiative, UBC 23. National Pharmacy Cardiovascular Council	1. Canadian Journal of Cardiology 2. American Journal of Cardiology 3. Circulation 4. Bandolier 5. Evidence-Based Medicine 6. Cardiovascular Review Reports 7. JAMA 8. ACP Journal Club

TABLE 2. Fugitive Search – Acupuncture in Addiction Treatment

Commercial databases	Web library catalogs (using Library of Congress or National Library of Medicine (MeSH) subject headings)	Internet peer-reviewed sites	Internet search engines
1. Cochrane Library 2. HSTAT (technology assessment guidelines) 3. HSRProj (NLM) 4. Dissertation Abstracts 5. Article1st (OCLC) 6. Papers1st (OCLC) – conferences and paper abstracts 7. TRIP database (evidence-based medicine) 8. Ebsco Academic Search 9. Ebsco Canadian MAS 10. Alcohol and Alcohol Problems Science Database (ETOH) 11. ACUBASE (France) 12. CRISP (Computer Retrieval of Information on Scientific Projects)	1. UBC Library Catalog 2. BC Ministry of Health Library Catalog 3. Canadian Institute of Scientific and Technical Information (CISTI) catalog 4. Belinda Database (Buckinghamshire Health Authority Library) 5. HealthPromis (UK web catalog of health promotion) 6. National Health Information Center – Health Information Resource Database 7. CHID Online (Combined Health Information Database) 8. WorldCat 9. GAO Web Catalog 10. COPAC 11. NLM Locator Plus	1. UK Academic Web Directory 2. UK Social Science Information Gateway 3. OMNI (Organising Medical Networked Information) 4. Medical Matrix 5. Health Communications Network 6. Global Health 7. Health Index 8. Medweb Public Health 9. Medscape	1. Northern Lights 2. Altavista 3. Excite
In-house databases	Directories	Organizations contacted	Journals hand-searched
1. In-House Catalog 2. Analytics Database	1. ECRI HealthCare Standards 2. UHC Technology Assessment Monitor	1. US National Institutes of Health, Office of Alternative Medicine (Also known as the National Center for Complementary Medicine) 2. Center for Complementary and Alternative Medicine 3. Center for Addiction and Alternative Medicine Research 4. The University of Texas Center for Alternative Medicine Research 5. Tzu Chi Institute for Complementary and Alternative medicine (BC) 6. National Acupuncture Detoxification Association 7. Richard and Linda Rosenthal Center for Complementary & Alternative Medicine – Rosenthal Center Directory of Databases 8. Research Council for Complementary Medicine 9. Centre for Addiction and Mental Health 10. The National Institute on Alcohol Abuse and Alcoholism 11. Centre for Addiction Studies (UK) 12. NIDA Organization 13. Canadian Centre on Substance Abuse 14. Foundation for Traditional Chinese Medicine (UK) 15. Hooper Detox Centre 16. National Academy of Acupuncture 17. BC Women's Hospital 18. Vancouver Richmond Health Board (Detox Committee) 19. The Acupuncture Association of BC 20. Evergreen Treatment Services 21. Ministry for Children and Family (BC) 22. Hessepin County Medical Center 23. Merle West Center for Medical Research 24. Lincoln Medical and Mental Health Center 25. Yosan University of Traditional Chinese Medicine	1. Alternative Medicine Journal (peer-reviewed) 2. Alternative Therapies in Health and Medicine (peer-reviewed) 3. Journal of Substance Abuse Treatment 4. JAMA 5. American Journal of Acupuncture 6. Journal of Alternative and Complementary Medicine 7. Lancet 8. BMJ

Directory Database of Research and Development Activities (ReaD)

Hiroyuki Sato,
Kenshou Nakagawa, and
Yoshiyuki Maeda

JST, Japan Science and Technology Corporation

INTRODUCTION

Japan Science and Technology Corporation (JST) has developed a Directory Database of Research and Development Activities (ReaD) since fiscal 1997, and offered test service of ReaD since August 1st in 1998.

Now, the Japanese government aims to build up a nation creating science and technologies and invests actively in various basic researches on the basis of a science and technology master plan. These researches are aimed at creating new industries by the private business sector using the research results. In order to use the results, information such as the content of research has to be offered and opened to the public. JST conducted ReaD to spread the national research information by the Internet.

Generally, to grasp the research contents, we ask the responsible institute or researcher to confirm the research theme as a basic information. For that purpose, JST constructed a directory of research institutes, researchers, research themes and research resources as the nucleus of contents of ReaD for users as fundamental information (guidance).

When we developed the system, we formulated three principles of design namely, (i) collecting the open technologies as much as possible (standardization of design), (ii) open-ended design of the system and (iii) user friendliness. These principles are inherent in all ReaD functions.

In this text, the concept of development, outline of the system, functions and directory of ReaD are described.

DEVELOPMENT CONCEPTION OF READ

1. Abstract of ReaD

An overview of ReaD is shown in Fig. 1. The nucleus of ReaD consists of three research functions (Category search, keyword search and Multi-link function) and four kinds of directory database (Institute, Subject, Researcher and Resource DDB) with cross-references to other JST contents. ReaD is under way to make possible cross-reference with other related contents. The search function and four kinds of directory were developed in fiscal 1997. Other contents and the total system were upgraded in fiscal 1998. The Internet was selected as the insemination under consideration of foreign countries.

When an information disclosure system is designed, a computer system is important as a receptacle, but collection of information, that is information resources, is more important. ReaD has two collection methods: one is a questionnaire by JST and the other is information entered by research institutes and researchers.

Moreover, JST will construct a system which offers all the internal research information and ReaD play a window of national information in cooperation with the National Center for Science Information System, etc..

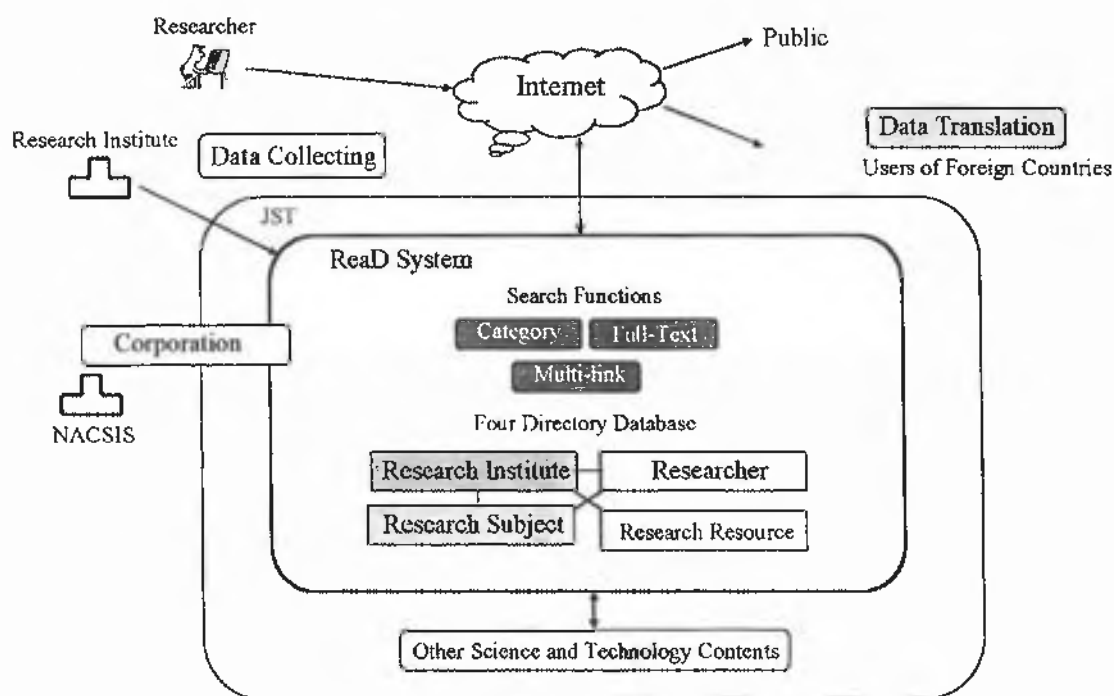


Figure 1 Concept of ReaD

2. Retrieval function

The most useful retrieval function is keyword (full-text) retrieval in the Internet. Although ReaD has keyword search, category search function is the main retrieval step, because user may retrieve names of research institutes and researchers and can search easily under a certain classification (category search) fixing the extent of the search breadth. Moreover, a part of them are added a squeezing function so as to offset a weak point of category search.

Keyword search displays various color of the appropriate character strings and changes the content of information at every directory and clarifies the retrieval results for user.

A multi-link function was developed as means of acquisition of all related information, because users might desire to obtain all information or research institutes and researchers without many retrievals.

3. Four kinds of directories

In the current printed medium, basic information are arranged under every research institute. The contents consist of research content (subject) of institute, researchers and offering service. JST distributes these data under various categories for use as one body by means of the Multi-link function. JST investigated the existing directory contents from the viewpoint of the kinds of information desired and which could be collected. Moreover, on the basis of know-how of JST and researches of the related organization, JST decided on the present four kinds of directories such as Institute, Subject, Researcher and Resource DDB.

4. Collection of information

Information is collected by a questionnaire on search institutes once a year at present. New entries and maintenance are carried out. Information on national organizations, such as National research institutes, collected by the Science and Technology Agency are used. Data on semigovernmental organizations and the public sector are collected by JST's questionnaire. These collection of data on data base is offered at any time by JST.

Information collection by the questionnaire is scheduled once a year, but updated every week. Accordingly, a representation of data correction and addition is accepted and updated at any time.

Information on universities is not directly gathered, but it will offer in cooperation with the National Center for Science Information Systems of Ministry of Education.

Although information is currently collected by questionnaire, JST has a plan to construct a system updated directly by research institutes and researchers by means of the following function.

5. Research entry site of information

Entry of information by researchers is very important in the conception of ReaD.

Although the entry format of Web is not special, a good format for entry of science and technology information and updating of it by researchers using this format are very important.

Although the Internet, a mass media next to newspaper and television, has home pages, data entry by researcher is not sufficient and the data is not used enough. Under these conditions, ReaD can offer the chance of information input by researchers. We are modifying the present update system of ReaD to make better system.

ABSTRACT OF SYSTEM

1. History of development

We designed data items at first to select data items of each directory and then clarified the mutual relation between them. The subject directory contains two categories: one is ordinary research under the theme of one institute and the other is a project pattern by many institutes and universities (one search has a hierarchical structure consisting of a large theme, medium and small themes). This pattern is increasing.

An original database were created by transferring the existing data to the new format and by questionnaire using the new format.

The system – data addition or updating, verification of contents and updating – completes one cycle. Data addition or updating was kept on the basis of annual questionnaire by JST and correction of data by researchers and JST. The contents are verified by a quality control officer at JST for integrity assurance. Data is updated once 1 week (Fig. 2).

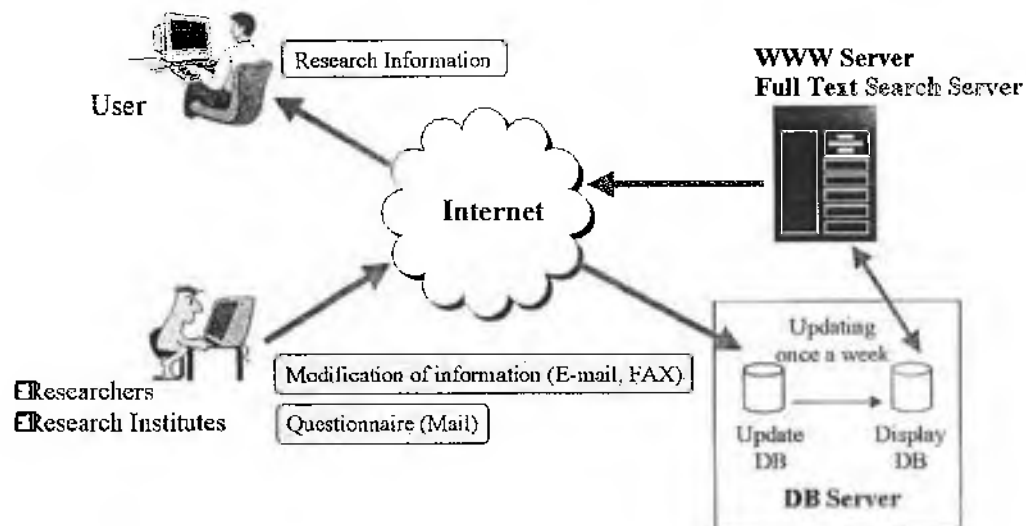


Figure 2 Image for ReaD system Operation

2. Hardware

The basic structure consists of two service systems, Service System 1 and 2, each made up of two servers such as WWW and full-text search processing server and a database server (Fig. 3). This structure can continue uninterrupted service during updating and in case of failures in either system. Moreover, an access load can be distributed to two systems for increasing further user convenience

All equipment are installed within the fire wall of JST network.

A database server in Service System 1 is WWW server for update processing via Web. Further, WWW of Service System 1 performs an electronic billboard function, statistics processing function, mailing list server function and output function of questionnaire results.

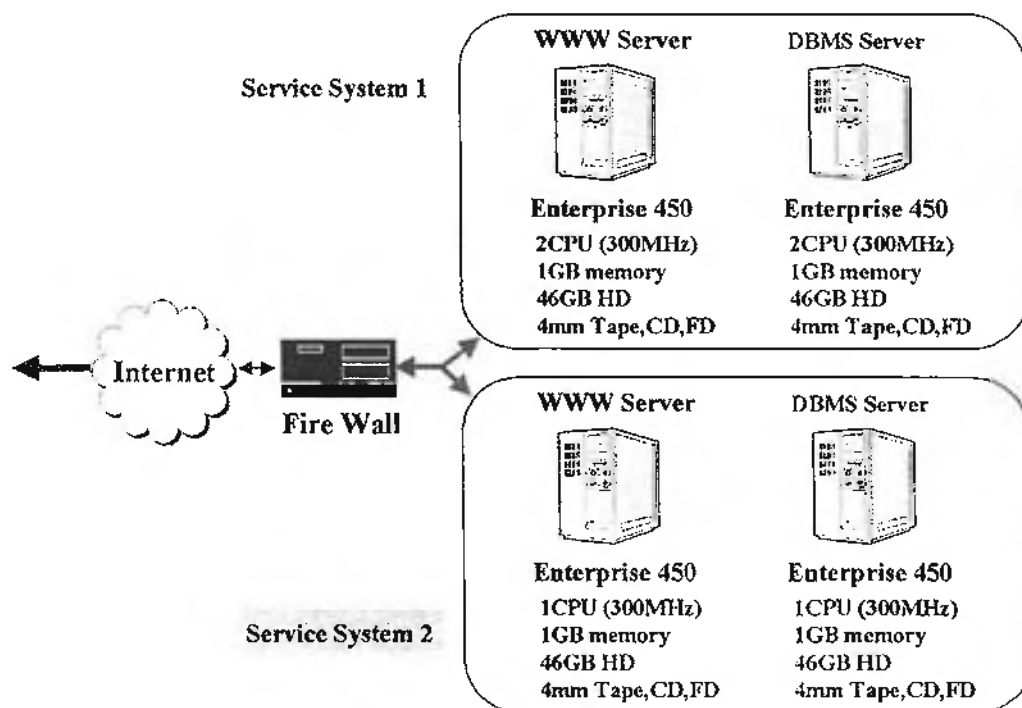


Figure 3 Read Hardware Structure

3. Software

The structure of software is shown in Fig. 4. Oracle Web Application Server 3.0 (WAS) is used as WWW Server. Oracle8 Enterprise Edition, Open Text and Livelink Spider are used as software of DataBase, Full-text Retrieval Engine and Robot Software, respectively.

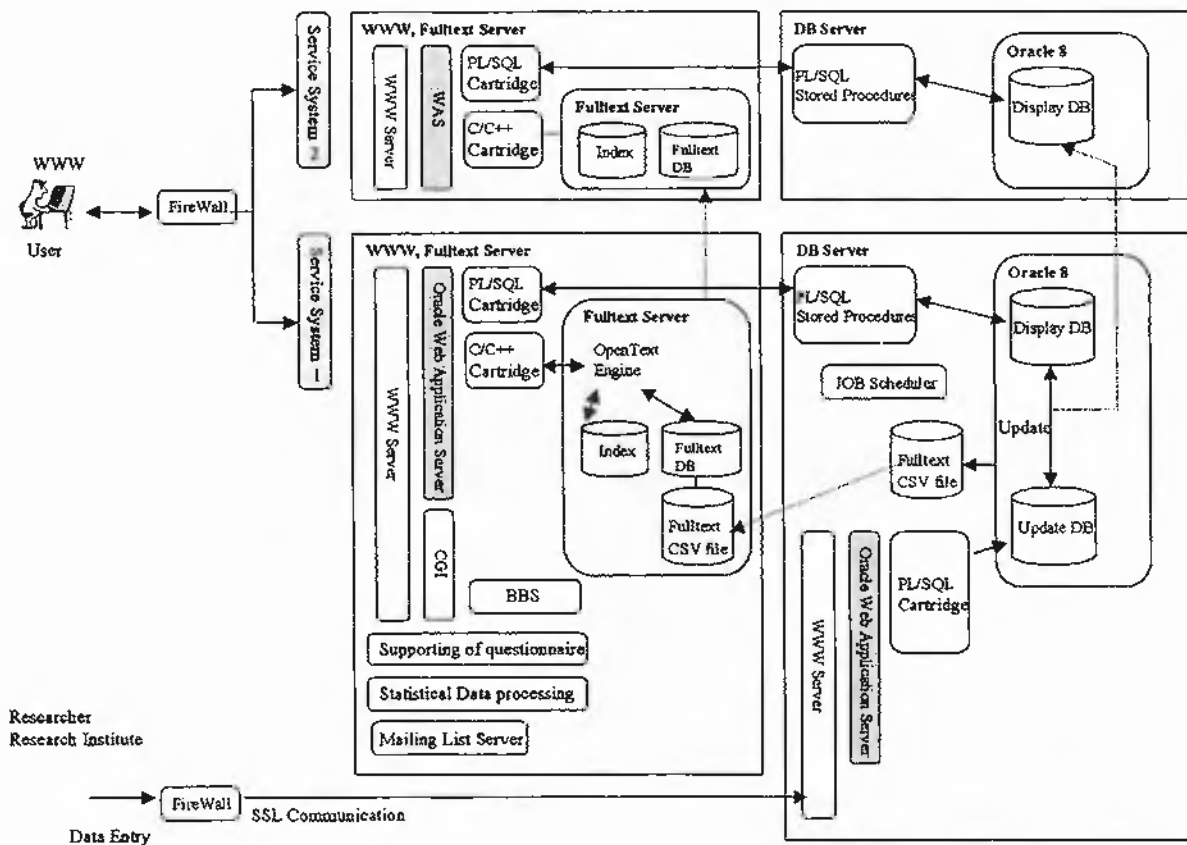


Figure 4 Software structure of ReaD

WAS consists of three parts: first is Web Listener as HTTP Server, second is Web Request Broken (WRB) equipment for HTTP Server API (Application Programming Interface) and third an external program named by cartridge (Fig. 5). A retrieval request entered by user's browser is accepted at once by Web Listener, then transferred to WRB for definite response to the cartridge and then the request can be processed by the selected cartridge. The retrieval results obtained by the above processes are returned to the user via the reverse route.

Cartridge automatically produces the number of process correspondences with the user's request while controlling load. The processing between cartridges is independent. Accordingly, error by one cartridge function does not affect other functions. Development of new Web applications means addition of new Cartridges. Use of cookies makes possible transaction management.

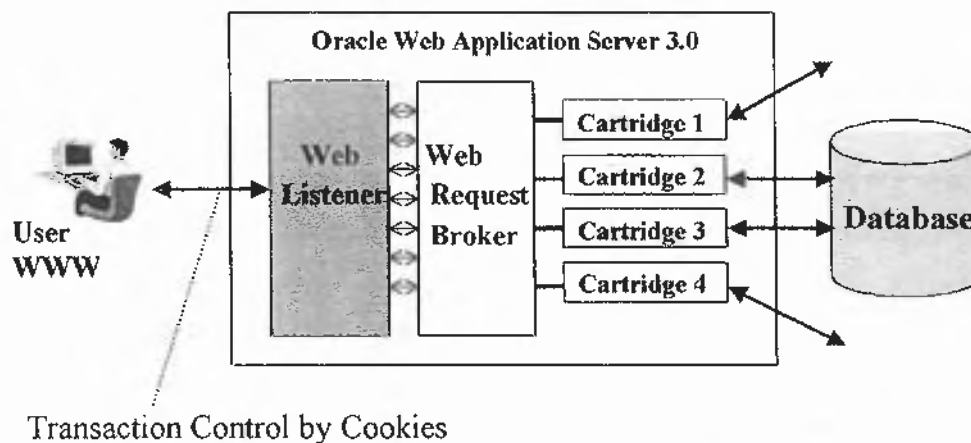


Figure 5 Link between WWW and DB

Next, a retrieval processing of ReaD is explained (Fig. 4). On the Category retrieval, a processing request from user's Web browser is accepted by WAS. When user requests "Announcement" or "What's New", HTML file is displayed directly to user. If user requires to access the directory database, the retrieval request is sent to database server via WAS, then the obtained retrieval results are displayed by HTML file after automatic reconfiguration. In the case of full-text search, the retrieval words and conditions entering into WWW Server are sent to full-text search engine via WAS, and then obtained results are displayed by HTML file. The retrieval requests of users are accepted by alternating acknowledge between Service 1 and Service 2 by means of the load distribution function of Fire Wall.

Updating processing can be done by WWW server on the Database Server of Service 1 as a special-purpose entry via Web. Reliability between user's browser and Server is increased by encryption using SSL (Secure Socket Layer) communication.

New data entered and update are placed on the update system database at once, and sent to update processing after verification of the contents by JST at week end. Subsequently, they are recorded on the database. Update processing of full-text system is carried out in the following order, the DB Server of Service 1 is updated and then CSV file (tag delimiter text file) for creation of index file of full-text search is swept. Then, DB System of Service 2 and full-text System of Service 2 are updated. The course of general update operation is terminated about half a day from the past records.

ABSTRACT OF FUNCTIONS

1. Retrieval function

There are two search methods of the directory type contents, one is a category search which the user follows the items of which are arranged by specific sorting and the other is a full-text search that the user can retrieve by use of any character string. ReaD offers these two search methods.

As category search, the user can obtain the objective information by sort items in each directory data (Fig. 6). ReaD offers "Squeezing search" function by which the user can squeeze the candidate on the way to retrieval (squeezing by the search system, research budget and research field in the research subject directory and squeezing by the research field in the researcher directory)

The buttons, showing link to DDB data and the related information, are set at the lowest part of the last display screen of each directory (Fig. 7). One of the important characteristics of ReaD is the relational links between 4DDBs named by a multi-link function. For example, user retrieves a research institute and then immediately can display its researcher names, themes and resources by means of it.

On the other hand, the full-text search function can be used by choosing DDB and input the retrieval words. Detailed retrieval using many retrieval words and the retrieval conditions such as AND and OR can be executed as well (Fig. 8).

With creating the index file of total Web server data in JST domain by using a robot software, the user can perform full-text search on them.

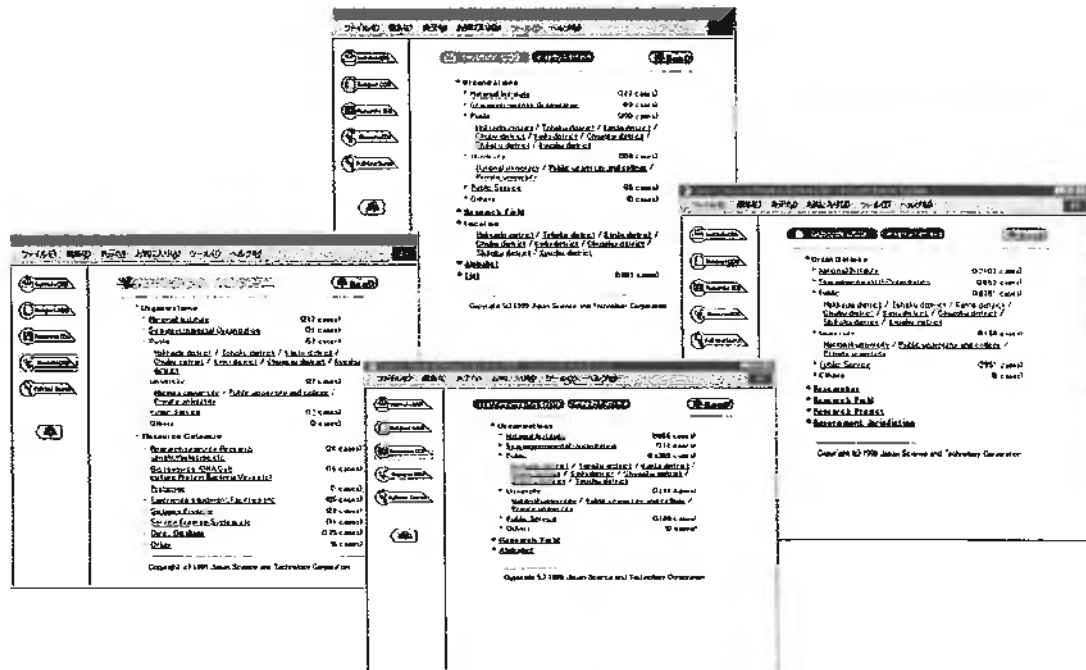


Figure 6 Read Category Search

2. Input data function

Input data to Read system is provided anywhere user can access the Internet and use a WWW browser. To use this service, users are requested to register user's ID and password issued by JST. Input function fixed the limit of data addition editing by creating three kinds of user categories, specifically, researcher, institute manager and JST manager. (Researcher can add and edit the researcher and research subject information, the manager of institute can do the above information and the research institute and research resource information and JST manager can carry out addition and editing of all data).

Read has also an output function of questionnaire results and a statistics processing function for managers. Both functions support the questionnaire work and analysis and control of information.

CONTENTS OF EACH DIRECTORY

The date is set automatically in all directories to express the date of input of the directory contents by the user. Moreover, it requires each research institute, researcher and JST data manager to pay attention to update.

1. Research institute directory

Table I shows the data items of research institute directory. The items explain about the institute such as its outline, history and purpose of activities. An address guide shows a map of the institute and the block plan of buildings. The items of research field, recent research results and department are used as an abstract of institute activities when user applies or utilizes it. The links are established between each directory page and home page of research institutes. It gives detailed information and can be used as link group.

Table I
Data Items of Research Institute DDB

1	Data Revision Date
2	Organizations
3	Address
4	Foundation Date
5	Telephone
6	FAX
7	History
8	Purpose of Activities
9	Main Activities
10	Representatives
11	Number of Staff
12	Number of Departments
13	Number of Attached Facilities
14	Area of Site
15	Area of buildings
16	Budget
17	Periodic publications
18	Attached facilities
19	Research Field
20	Departments
21	Recent Research Results
22	Number of Patents
23	Guidance of Address
24	Traffic

2. Research subject directory

Data items of the research subject directory are shown in Table II. The items consists of data on the name of researcher, accomplishment, completion/in progress and research field. The subject directory characterized to cope with a large research theme (main theme) with some smaller themes, for example, Exploratory Research for Advanced Technology (ERATO) of JST. In this case, the item of project, representative organization of project and representatives are used. It makes possible to retrieve research of a specific research system. Information about research results and evaluation of research supported by national fund are entered into it after due consideration that these research results are asked to be opened to the public. The research evaluation has no data open to the public, but we will develop a directory that you can look through the subject from beginning to the accomplishment, prize, etc., by entering data from the evaluation reports.

Table II
Data Items of Research Subject DDB

1	Data Revision Date
2	Research Subject
3	Research Organization
4	Departments
5	Researcher
6	Project
7	Representative Organization of Project
8	Representative
9	E-mail
10	Research Budget
11	Research Duration
12	Finish / Undergoing
13	Outline of Research
14	Research Field
15	Keywords for Research
16	Research Project
17	Government
18	Accomplishment
19	Prize etc.

3. Researcher directory

The data items of researcher directory are shown in Table III. There are many items about researcher personally, research contents and the past records which are not the required items but they are provided upon the researcher's discretion. In four directories, Researcher directory contains personal information, that is private information, so that only data input by researcher himself is offered and opened to the public including with data of the directory.

If the researcher desires to provide a lot of information, it is possible without any limitation on space. In this case, ReaD runs a set of operation that is repeated regularly in ten papers and books. When item of paper, book and patent stored many data, these items in Researcher directory become very useful information.

Home page URL of researcher, present research theme and free comments are offered for researcher to present himself. It is recommended to use them as much as possible.

Table III
Data Items of Researcher DDB

1	Data Revision Date
2	Name of Researcher
3	Organizations
4	Departments
5	Title
6	Birth Year
7	Graduate School
8	University
9	Degree
10	Mailing Address
11	Telephone
12	FAX
13	E-mail
14	Personal Homepage URL
15	Societies
16	Research Experiences
17	Career (Professional)
18	Research Field
19	Keywords for Research
20	Research Theme
21	Prize
22	Papers
23	Books
24	Patents
25	Availability (Books, Data, etc.)

4. Research resource directory

Finally, the items of research resource directory are shown in Table IV. Research resources in this article is a resource that is material and immaterial resources (hardware, software database, various kinds of systems and service, etc.) maintained by organizations and a outsider can use them by proceeding through the necessary formalities.

The items of the directory contain a list of organizations and contact persons who manage the resources, research subject category, resource information, condition for use and utilization.

Table IV

Data Items of Research Resource DDB

1	Data Revision Date
2	Research Resources
3	Mailing Addresss
3-1	Organizations
3-2	Departments
3-3	Contact Person
3-4	Address
3-5	Telephone
3-6	FAX
3-7	E-mail
4	Research Subject Category
5	Research Resource Information
6	Samples
7	Condition for Use
8	Utilization (Application)
9	Related Information
10	Formality for Use

CONCLUSION

At the end of introduction of ReaD, the objects of ReaD are explained.

(i) ReaD aims at one stop Web site of science and technology research information. JST is developing the necessary contents, for example, a symposium guidance directory and a science technology link group and user friendly retrieval functions.

(ii) ReaD performs an entry point to open research information of the nation, etc. Various kinds of research information for example, research results information without the present directories, are created and linked with ReaD and offered as an integrated database. ReaD will construct the system with research information of nation and universities in cooperation with the National Center for Science Information Systems.

(iii) ReaD aims at an entry site of research information by the research institutes and researcher. As indicated above, the entry site of information is very important for security of information entered by researcher himself and for collecting science and technology information in future. JST is going to develop the necessary functions which researcher himself can enter information and to spread use of the site.

Note: Oracle8 is a trademark of Oracle Corporation.

Open Text and Livelink Spider are trademarks of Open Text Corporation.

Towards integration of information sources on grey literature. A case study

Rosa Di Cesare, Consiglio Nazionale delle Ricerche (CNR) - Biblioteca Centrale -, P.le Aldo Moro, 7 Rome - Italy; dicesare@bice.cnr.rm.it
Gianni Lazzari, Biblioteca della Camera dei Deputati, Via del Seminario, 76 Rome - Italy; lazzari@camera.it

ABSTRACT

The institutional goals and the organisational contexts of GL producers, determine the level and the process of production and diffusion. Furthermore GL producers don't bring all the same "dowry": they have got neither the same resources, nor the same information policy. Inspired by such considerations the Italian reference centre of SIGLE (System of Information on Grey Literature in Europe), has based its activity on the Cupertino with GL producers.

To achieve this aim the Centre has evolved from a merely receptive role, consisting of pure and simple acquisition and bibliographic treatment of documents, to an active role, in order to control the very process of document production by exerting a consulting function on bibliographic aspects. In other words is the Centre that places itself at disposal of producers to solve all the problems connected with their participation in SIGLE database.

As an example of this policy the paper describes the interaction between the SIGLE Italian reference centre and the Library of Italian Chamber of Deputies which produces the database LEGO, which stands for Grey Literature Online, containing non commercial, official and administrative documents, produced and/or acquired by the Library since 1993. The consulting activity developed by the centre has allowed a fast and efficient acquisition of the document stored in LEGO.

Finally the paper describes the type of documents in LEGO, the selection criteria and the issues connected with information integration.

1. Introduction

The development of information technologies in the information and documentation field and the concomitant effects on working processes and products (electronic catalogues, interconnections, information sources integration) allow, albeit potentially, the effective distribution of the information sources and therefore, pave the way for change from the conception of documentary resources mainly based on "possession" to a conception based on "access" (8).

But if the improvement of the information and documentation technologies has implemented the capacity of producing and transferring information, on the other hand it has only partially solved the problems concerned with the management of data coming from different archives (or stored in different ways).

The problem is that the distinct information sources differ (inside) not only in data structure and in the treatment of information but also in organisational and technical models (2).

Integrating information is therefore a complex activity as it involves some basic aspects of work organisation. Different procedures and different "ways of working" and of treating information have to be integrated with technologies. An information system should solve these differences by identifying, both the basic requirements for integration (for example document types, management and bibliographic treatment), and the means of harmonisation (use of different systems for cataloguing and classification).

Information source integration is not new to the world of information and documentation. A first important way of integrating information sources has been in fact the realisation of collective paper catalogues, the evolution of which is testified by OPAC (6, 13).

An automatic transposition of the above experiences on the GL field is quite difficult, given the well known characteristics of GL itself: above all diversification of producers, fragmentation of information sources and the lack of bibliographical control. Moreover, even if the producers make a bibliographical control of these documents, they may do so in very different ways.

Inspired by such considerations, the SIGLE Italian reference centre, has based its activity on co-operation with GL producers, and with one in particular, namely, the Library of the Chamber of Deputies, which produces the LEGO (On line Grey Literature) database where official and administrative documents have been collected, produced and/or acquired by the Library since 1993.

In this work the introduction of the experience of the co-operation above is preceded by some remarks on GL collection and management. In section 2. some features relating to particular producers and those related to the activity of the SIGLE Italian reference centre are considered. In the last section some considerations of the experience are given.

2. Some remarks on GL collection and management in Italy

Generally speaking, in Italy the fragmentation of Libraries and Documentation Centre, apart from scientific and technical institutions (up to cultural foundations) has partially delayed the realisation of the co-operative information system. As regards GL the SBN (Italian National Library System) has considered other national and European GL databases, by introducing some bibliographic details peculiar to SIGLE standards. However, the input of GL documents to the SBN database is, up to now, extremely limited. It is probably due to the lack of specialist support, especially during the phase of selection and collection of documents.

Further isolated initiatives that are to be found are compartmentalised in nature and deal exclusively with specific document types such Doctorate theses which have only been dealt with in a specific publication in the Italian National Bibliography since 1995, or with specific disciplines. The key public research bodies handle scientific and technological GL while legal and administrative GL is handled by the Library of Chamber of Deputies (1).

2.1 GL producers: some characteristics of GL management

Turning to individual GL producers, the situation is varied. In 1996, the first survey carried out by the SIGLE Italian reference centre on individual producers' organisations revealed a

situation that was far from uniform. Italian GL producers adopt different approaches to the handling of documents and to the choice of appropriate technology.

The survey shows that acquisition of GL documents takes place via

- *a direct channel* between the "historical" producing bodies and the Italian reference centre (50% of the cases);
- *a direct channel* of legal deposit between the producing bodies and the Central Library of the National Research Council (CNR) to which the SIGLE Italian reference centre has selected access (25%);
- *a direct channel* between the "recent" producer and Italian reference centre, set up for the bottom up release of data for users of the Central Library and for librarians at annual training courses run by the Library itself (15%);
- *"informal" channels* relying on "informal" contacts mainly between academics, library users, librarians and CNR documentalists (3).

Further information on GL producers came to light on analysis of factors such as document layout and the organisation of bibliographic elements of cover page.

This information breaks down to cover three categories:

- Organisations that have a highly organised information network along with in house editing facilities and/or library, both internal and external GL documents are collected and subjected to bibliographical control;
- In the middle ground there are organisations that conserve in house GL making it the responsibility of a librarian or head of department. The latter nearly always applies personal criteria regarding the conservation and handling of documents, with scarce diffusion resulting. The SIGLE Italian reference centre has offered these organisations support facilities for bibliographic control;
- At the other extreme there are organisations that merely archive their GL in "bureaucratic" fashion, failing to carry out any bibliographic classification of the material. For this group the Italian reference centre prepares training schemes and provides information.

In the comparison between GL producing offices and organisations one common factor emerges. Generally speaking GL is not subject to efficient bibliographical control nor is utilisation monitored. It should be pointed out that some organisations possess advanced documentation systems in qualified hands, with access to quality data treatment systems in an environment with little interest in specific GL projects, especially when these require standardisation, extra personnel and funding. Having said this, there has been some recent interest in integrated acquisition of GL data.

2.2 The SIGLE Italian reference centre

Reviewing its activities, the Centre has had to embrace the wide variety of individual situations and general characteristics pertaining to GL information sources and management that has been outlined above. Key considerations taken into account are: 1) information sources and GL producers differ not only from an editorial and technological point of view but also in the bibliographic management of their GL documents; 2) the institutional goals and the organisational context of GL producers determine the level and the process of production and diffusion. Moreover, GL producers don't all bring the same "dowry", having neither the same resources nor identical information policy.

In the light of the above and in conjunction with its institutional functions, the centre has sought to concentrate on research and information sources as well as GL producers, trying to assure their participation in the SIGLE database. This work has itself given rise to new initiatives while at the same time consolidating and rethinking the traditional parameters of databases.

The centre's chief commitment has been to research into new producing organisations and to provide support for those already in existence, participating in the database, by promoting enhanced bibliographic management of GL documents in line with commonly agreed standards.

Pursuance of these objectives has been hampered by two elements: one to do with the inherent properties of GL, and the need to know the specific grey routes taken in the production and diffusion of GL documents, frequently limited to a specific production environment (a commission, a project etc.); the other due to organisational and functional shortcomings in libraries and heeldragging on the behalf of GL producers (both individual and organisations) who fail to appreciate the importance of documents produced in working groups and research projects and effectively block acquisition and circulation in the scientific community.

Given the situation, the Centre has emerged both in practice and in institutional terms as the national reference point for GL. In order to maximise producer involvement the Centre has evolved from a merely receptive role, consisting of pure and simple acquisition and bibliographic treatment of documents, to an active role, in order to analyse the very process of document production by exerting a consulting function when necessary and/or requested, on bibliographic aspects with a view to integration of the different information sources. The assistance and consultation services provided by the centre deal with both editorial presentation of documents (whether or not to bring out a new series, project differentiation according to content etc.) and with bibliographic description of document (attribution of catalogue numbers, series titles, standardisation of the producing organisation etc.).

In other words it is the centre that places itself at the disposal of producers in order to solve problems arising from their participation in the SIGLE database.

The above mentioned evolution in the centre's activity has been made possible by the strategies adopted by EAGLE (European Association of Grey Literature Exploitation) which produces and runs the SIGLE database. These strategies have permitted a more flexible handling of SIGLE standard both in terms of bibliographic format and bibliographic description of documents. (Implementation of new software that is directly operated by the various national reference centre of the countries adhering to the SIGLE system, simplification of the bibliographic level of description (monographic or serial) and

the capacity to acquire documents with the title in the original language so long as key terms are given in English.)

3. Toward integration between the SIGLE Italian reference centre and the Library of the Chamber of Deputies

The SIGLE Italian reference centre has opted for a form of Cupertino and integration of GL information resources with one selected producer, namely the Library of Chamber of Deputies. Four factors prompted this choice: 1) the library itself having set up a specific GL database; 2) its national role and institutional standing in the field; 3) the availability of GL documents produced by key organisations and their inclusion in the LEGO database; 4) the library having personnel skilled in the handling of such documents.

3.1 The LEGO (On Line Grey Literature) database

The LEGO database came into being in 1993 and currently holds some 10,000 bibliographic references with an annual growth of 1,700 -1,800 new inputs. It is a parliamentary grey literature database for material emanating from the Chamber of Deputies and the Senate or coming into their possession from governmental bodies, professional associations, universities, Regions or local authorities, research organisations etc. It provides day to day back up to parliament. The most striking characteristics of this unusual database is its dual function. The documentation that is acquired is produced both by parliament itself, and on the whole, does not appear in Parliamentary bills, as well as by data reaching parliament in the carrying out of its legislative and monitoring activity, above all in preliminary hearings of permanent commission.

Clearly such data is of great institutional and political importance. The availability to the public of this data is an Italian parliamentary commitment towards the divulgation of the legislative procedure, its inspection and monitoring activities, via open access to the texts at the basis of political decisions.

LEGO provides access to hidden channels of information that are often unknown and which vary in type and source. These are pooled in a single reference structure with a single reference system. This is to the advantage not only of citizens but also of the administration itself. The more sources are available, the greater the benefits.

Clearly the database seeks to meet the information needs of its primary users in the library and the Chamber. To be precise these users are Deputies, political groups, Parliamentary Commissions, service providers and offices. A co-ordinated and indexed collection of documents that are essential for the carrying out of specific functions is made available.

Documents are grouped in two categories: 1) internal; 2) external, and this reflects both means of acquisition and contents. Both categories are further subdivided according to the needs of the complex institutional activities of the Chamber. There is almost total bibliographic control of the first category and this permits a range of document products, normally organised in series. Nomenclature reflects the mainly grey channels followed by this literature that is produced in support of legislative activity at all levels. (Documentation and research, Decree dossiers, Actions dossiers).

The second category, those documents that are sent to the Chamber is extremely disparate reflecting the range of producers. The control carried out on these documents by the library is to date only partial and is beset with difficulties.

It should be noted that "accompanying slips" are attached to parliamentary documents enabling the successive steps of the legislation to be followed (Presentation, Discussion, Approval) and the elaboration of interesting statistics on parliamentary proceedings (12).

1.1. Cataloguing and structure of LEGO documents

Documents are recorded in the Library's DOBIS/LIBIS handling system in accordance with ISBD standards. The system has undergone various simplifications when it was felt that rigidity did not permit exhaustive bibliographic description of GL documents. Indexing is carried out by the TESEO Thesaurus, which is an instrument set up by the Senate for parliamentary documents based on Universal Decimal Classification (UDC).

Elements in the DOBIS index are inserted into the LEGO database. Some specific fields of selection and research in the bibliographical headings are of particular interest. Consider the *Tipdoc* field (document typology) and the *Freq* (frequency) field, which are used respectively to select documents according to the means of acquisition and the frequency of despatch of periodical reports. In the main body of bibliographic description we find the *Materia* (subject matter) field containing the TESEO thesaurus terms, the field *Fonte normativa* (source of legislation), where the standards of the legislation or the essential elements of Bills are recorded. The field *Iter* (course taken) records procedural information on documents named in the Chamber and assigned to a Commission.

The co-operation with the SIGLE Italian reference centre has benefited from the streamlined input of LEGO data into SIGLE thanks to the utilisation of TESEO. This is done by means of UDC subject categories in English, which has greatly facilitated the indexing of LEGO documents with SIGLE subject codes (10). Currently, SIGLE subject codes are added to TESEO markers on implementation of the LEGO database. This permits greater harmonisation between the documents that are inserted and the SIGLE database. The net result polarises into LEGO autonomy, and at the same time, LEGO compatibility with SIGLE. (The latter benefits by gaining enhanced value of its documentation with a minimum of extra document description work.)

3.3 LEGO documents in SIGLE

The following is an outline of the phases in the Cupertino. During its work use was made of an "agreement" defining techniques for the eventual overall integration of information sources (4).

- *Selection.* During this first phase LEGO selected only a part of the *Internal documents* that had already been subject to bibliographic control by the Library of the Chamber of Deputies. It was decided to acquire only those documents whose contents were of interest to a wider readership (Actions dossiers, Documentation and research etc.). The external documents selected were largely Government reports or those from the State Audit Court.

- *Bibliographic description.* Integration was made easier by the producer having adopted a flexible application of the ISBD standards. A case in point is the classification of LEGO registered documents according to document typology.
- *Format.* The SIGLE Italian reference centre provided a simplified conversion format for the FIBRE programme (5.) The format is based on a registration card organised according to document typology. Data travels from the Library of Chamber of Deputies to the centre by e-mail in standards SIGLE format.
- *Indexing.* During this first phase the Library of Chamber of Deputies uses SIGLE subject codes to index the documents. More detail on content is given by assigning UDC markers in the keywords field. An automatic conversion from UDC to SIGLE subject codes is being studied.
- *Titles.* The recent SIGLE decision to accept documents with titles in original language, as long as this is supported with key terms in English has enabled acquisition of LEGO documents.
- *Availability.* The agreement attributes an equal position to the two signatories that share responsibility for document delivery. This means that documents can be requested from either structure.

Up to now he has concentrated on improvements to procedure. We stress the benefits of having eliminated the transmission of documents by the producer to the centre. This has saved costs during registration and physical sorting of documents for all concerned, particularly for the centre. Producer harmonisation with SIGLE standards has freed resources and slashed SIGLE document acquisition costs.

4. Closing notes

Starting just after World War II, the CNR Central Library has collected GL from important national and international scientific, public and private institutions. This activity was extended and developed with adherence to EAGLE. Furthermore, under terms of legal deposit, the Library itself receives Italian technical/scientific publications. This legal right has habitually included GL documents.

The Centre's activity can be divided into two complementary phases. The first broadly coincides with the setting up of the SIGLE Italian reference centre at the Central Library of the CNR in 1985 (membership of EAGLE goes back to the same year), and continues until 1995.

During this first phase, the Centre's activity has concentrated on elaboration and management of Italian data (acquired from the Library as indicate in 2.1. paragraph). This data is periodically sent to the SIGLE database. Analysis of GL producer profiles was not carried out at this stage.

In the second phase the SIGLE Italian reference centre has implemented tailor made measures for individual producers in the drive for implementation of information resources. Producers' needs were analysed and care was taken to minimise interference with already elaborated technical-bibliographic procedure for GL documents.

This activity can be broken down as follows:

- *First level:* assistance to producers on grey information enhancement and the initiatives to be adopted. It should be borne in mind that rapid integration is out of question;
- *Second level:* aimed at producers who have already spontaneously adopted GL management measures. In these cases, the centre proposes an equal partnership aiming at the development of common strategy in GL management. This goes for theoretical considerations and/or the drawing up of guidelines. At this level, where the emphasis is no longer on urgency, more stable forms of collaboration can be established (common actions, "agreements"), that help towards integration of information sources.

Measures intended to involve producers must also:

Moderate the scale of participation: producer's involvement has to be calculated so as not to overload the SIGLE Italian centre with requests for assistance. Wider participation causes problems that are not only technical (different systems and instruments), but also problems of communication between the Centre and the various producers and between the producers themselves. In an integrated system, presumably producers must interact with one another and not only with the centre (e.g. the exchange of documents).

Monitor the producers and publicise the process of integration of resources: in the SIGLE Italian reference centre's current phase of activity, the monitoring is needed to have an idea of the progress being made (producer participation levels): data flow is still prevalently unilateral (from an individual producer to the centre).

As a result of this collaboration ground is being covered on the selection of individual producers to act as possible "hubs" in the widest possible network.

The aim is to publicise GL producers and their information resources through the Centre, and includes the benefits of elements of comparison arising from having to overcome common problems using analogous treatment procedure for grey information.

Monitoring can become a viewing glass for the harmonisation of procedure and for the transparency and identity of individual producers.

Further initiatives regarding integration could derive from "publicity" in appropriate channels (e.g. the Italian Libraries Association - Bulletin) on the resource integration initiatives that are underway.

It cannot be overlooked that in this potential increase in the multiplicity of communications the SIGLE Italian reference centre will have to find new partners and will need to reinforce its role of sorting and filtering grey information as well as that of putting producers, users and resources in touch. (When possible steering demand for grey information directly to where it is produced).

These new steps in the integration of information resources will involve the Centre in a *supervisory role* under contractual obligation with the different subjects involved, as in the case of the "agreement" with the Chamber of Deputies.

In order for integration policy to be successful, a communication culture is required. This should be understood in the etymological sense of the term i.e. pooling of objectives, results, criteria, procedure, and all else required for the development of information resources.

To sum up two issues arise. One is that the application of a standard model at best slows things, at worst making it impossible to proceed towards the objective. It is our experience that imposing a standard frightens off GL producers. The adoption of any standard requires technical, organisational and often cultural change. The other one is that the centralised organisation should be replaced since it is incompatible with the new scenario. The SIGLE Italian reference centre as a problem interpreter and shooter can, by means of integration, set up a beneficial circuit between producers, offering the end user an innovative "product" in terms of quality and quantity of GL information sources.

References

- (1) Alberani, V.; De Castro Pietrangeli, P.; Minutoli, D. (1996) La letteratura grigia. 2. Convegno Nazionale, Rome, 20-21 Maggio 1996, Istituto Superiore di Sanita'.
- (2) Colin J. (1998) Electronic technology and its impact on libraries, *Journal of librarianship and information science*, 30 (1) 7-24.
- (3) Di Cesare, R. (1996) Il Centro Nazionale di Riferimento SIGLE nella gestione della letteratura grigia prodotta in Italia: ruolo e prospettive. In: La letteratura grigia. 2. Convegno Nazionale, Rome, 20-21 Maggio 1996, Istituto Superiore di Sanita'.
- (4) Di Cesare, R.; Lazzari, T. (1999) La base di dati LEGO della Biblioteca della Camera dei Deputati in SIGLE, *AIB Notizie*, 11(5):1-13.
- (5) FIBRE (Friendly Input of Bibliographic Records) produced by INIS/IAEA adapted to SIGLE standard
- (6) Gatti, G. (1997) Macchine celibi? *Biblioteche Oggi*, 15(6) 6-21
Scritti in occasione del Centenario della FID (ed.) Augusta Maria Paci, Franco Angeli.
- (7) Jamshid B. (1997) The evolving OPAC, *Cataloguing and Classification Quarterly*, 1997, 24 (1/2), 163-85.
- (8) Poppi, G. (1996) *Gestione di risorse integrate per il recupero dell'informazione e la fornitura del documento: esperienze italiane*, in La Documentazione in Italia. Scritti in occasione del Centenario della FID (ed.) Augusta Maria Paci, Franco Angeli.
- (9) Rutstein, M.; Demiller, A.L.; Fuseler, E.A. (1993) Ownership versus access, *Advances in Librarianship*, v. 17.
- (10) SIGLE Manual. Part 2: Subject category list, European Association for Grey Literature Exploitation, 1999.
- (11) Timoty Sprehe, J. (1999) Government information: from inaccessibility to your desktop and back again, *Journal of the American Society for Information Science*, 50(4), 340-345.
- (12) Venturini, F. (1995) La letteratura grigia parlamentare. In margine a una nuova banca dati, *Bollettino AIB*, 35(3), 321-342.
- (13) Weston, P.G. (1996) *Le tecnologie dell'informazione*, in La Documentazione in Italia. Scritti in occasione del Centenario della FID (ed.) Augusta Maria Paci, Franco Angeli.
- (14) Zahiruddin, K. (1998) System migration: challenges for libraries in the Arabian Gulf region, *The Electronic Library* 16 (3), 171-174.

INTERNET USE

Research of Grey Literature by teachers-researchers at the University of Constantine

SEMRA HALIMA

Institut de Bibliothéconomie
Université de Constantine

To have access to scientific and technical information, the university teachers-researchers always tend to use the documentary structures of the university of Constantine.

Evidently, they also go -or are orientated- to other information sources out of the university : public libraries, archive centers, documentation centers in companies..., according to their information needs.

The recent integration of Internet at the university of Constantine has incited us to lead a survey that will allow us to know whether this new mode of access to information will have a distinguishing place in the teachers-researchers' habits, and whether grey literature is taken into account in electronic research.

1. INTERNET AT THE UNIVERSITY OF CONSTANTINE

The university of Constantine was born as a result of the reform of the Higher Education System which took place in 1971 and conferred to this institution the statute of multi-disciplinary university in the whole eastern region of Algeria.

Its population does not stop increasing and has reached this year (1999) about 1500 university teachers and about 35000 students.

The documentary resources spread all over certain structures such as :

- The university library which represents the most important structure.
- The institutes' libraries (22).

There is a great interconnection between these units so that they may be able to answer their users' needs in terms of studies, teaching, activities and research.

The policy of establishing a national network of scientific and technical information has led the (CERIST)* to provide the structure of Higher Education and Research with the services offered by the « network of networks ».

The integration of Internet at the university of Constantine took place in november 7th 1998 within the university library. This service has got a specialist telephone line connected to the CERIST.

The exploitation of Internet by the teachers-researchers requires a pre-registration due to the limited available places. The exploitation is however free of charge and offers :

- The bibliographical research.
- The electronic mail.
- The file-transfer.

Now, 180 consultation sets have been provided at the disposal of users and spread over 5 university campuses.

* Centre de Recherche sur l'Information Scientifique et Technique (Alger)

2. THE SURVEY

- a. Our choice has been focused on the university library, the first location where Internet has been integrated – November 7th, 1998 within a room (the Internet Room) that is equipped with 16 consultation sets.

- b. How was the survey carried out?

A questionnaire was administered to the persons who were using Internet in two periods of time:

- At the moment of the integration of Internet, that is to say November-December 1998.
- Six months after its integration: April 1999.

With few questions that were simple and concise, the majority of the teachers were able to answer the questionnaire the day of its distribution. Few, however, handed it in later. 101 teachers answered the questionnaire.

- c. The objectives of the survey

Our analysis intends to clarify certain questions to which we would like to find an answer:

- Who are the Internet users?
- What are their objectives in using Internet?
- What do they think of Internet? (advantages/disadvantages)
- How do they conceive the future of this technological tool?
- Does grey literature occupy a place in their electronic research?

Our hypothesis is based on the assumption that the use of Internet varies according to the subject.

3. THE RESULTS OF THE SURVEY

3.1. Characteristics of the population studied

Table 1: position of the population studied

Position	Answers
Professor	7
"Maître de Conférence"	12
"Chargé de Cours"	21
"Maître Assistant"	19
"Assistant"	38
Engineer	4
Total	101

The table reveals that the highest number of answers comes from the "Assistants". Nevertheless, when we observe the number of answers in terms of the global population – by position – we notice that the most important proportion of answers derives from both the extreme positions: Professor/Assistant. This is illustrated by the following table:

Table 2: percentages of answers by position

Position	% of answers	Total of population by position
Professor	9.2	76
"Maître de Conférence"	8.4	142
"Chargé de Cours"	3.5	588
"Maître Assistant"	4.2	452
"Assistant"	16	235

3.2 The Internet use- objectives of uses

a. The great majority of teachers use Internet with research purposes (90 answers out of 142), most of the time during preparation of a thesis (Master or Ph.D.) – an answer given evidently by the "Chargé de Cours", the "Maître-Assistant" and the "Assistant".

Some times – mainly – the Professors and the "Maître de Conférence" have recourse to tele loading of articles or preprints related to their field of research.

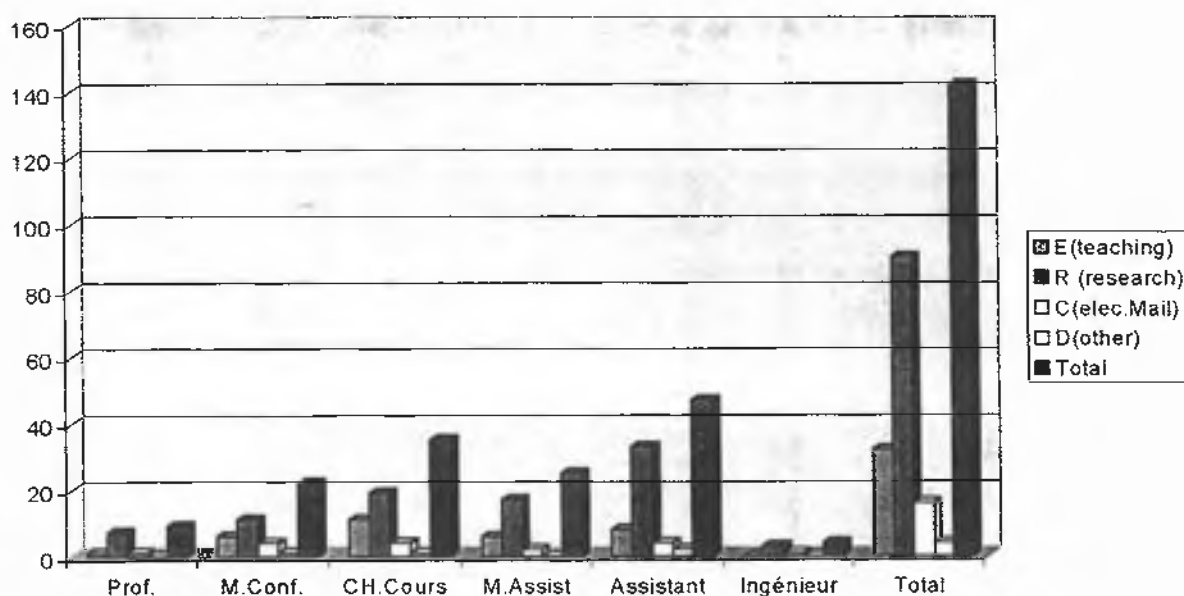
b. The second objective concerns teaching, Internet is used according to teachers, especially the "Chargés de Cours" and the "Maîtres –Assistants" for the preparation of their lectures.

c. Thirdly, the exploitatin of Internet is linked to the rapid exchange of information thanks to the Electronic Mail: an information means that cannot be neglected, but still for the time being less representative. We may say that we need time to be familiarized with such up- to – date tool.

d. From the other possibilities offered by Internet that attract the teachers' attention, we may cite: the press, entertainment- these cultural usage remain, however, of little use.

Thus, the exploitation of Internet concerns in priority the access to bibliographical references in the field of teaching and research.

Objectives of uses of Internet



3.3. The appreciation of Internet

a. Advantages

Most of the teachers agree to recognize that Internet has got advantages thanks to:

- The rapid access to information.
- The economy of time.
- The significant size of information available.
- The possibility to communicate with the whole world.

These parameters have led people to assume that this tool cannot be compared to anything else.

b. Disadvantages

Many teachers have thought about major disadvantage related to the time allocated to the consultation of Internet which seems to be not sufficient – a thing which is “frustrating” for certain questioners.

Other teachers criticize the slow process in getting the answers and the tele-loading which is not all the time possible.

Certain teachers underline the tremendous size of information – this raises the problem of saturation and over – consumption.

Few have dealt with tiresome as a result of Internet use.

We should mention that certain persons did not notice any disadvantage related to the use of this tool.

c. Suggestions

This tool is so appreciated that most of the teachers claim its valorization:

- By integrating Internet in all the institutes (institutes' libraries) – that is to say its extension to all university campuses.
- By providing more time for exploitation – at night if possible –
- By training a staff who may help the users.

This raises here the problem of redefining the role of the professional of information.

3.4. The place of grey literature in documentary research

The teachers - researchers seem to use various types of documents – but certain resources are largely more used than others.

a. Published articles

This type of document is the attraction of all: it comes in the first position with 95 persons out of 101. With a percentage of 41.30 %, it is then a priority.

b. Papers read in conferences

The recourse to such documents is important, especially for the teachers who occupy the highest academic positions : Professor, « Maître de Conférence », « Chargés de Cours ».

c. Books and theses

On the third position come these types of documents sought by means of Internet.

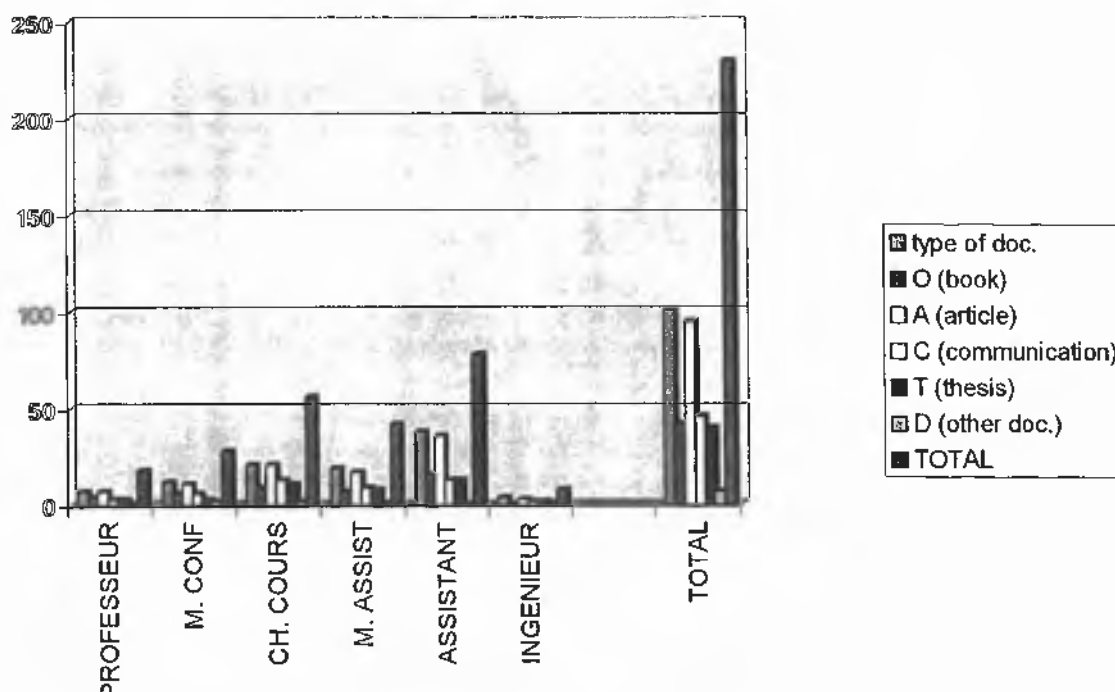
With a slight preference of one document to another, these documents seem to be used mostly by the “Chargés de Cours” and the “Maître-Assistant”, whereas the books show more attraction again among the teachers who occupy the extreme position: Professor, “Maître de Conférence” and “Assistants”.

d. Other documents

Among the documents that are mostly mentioned, we cite:

- Reports
- Softwares

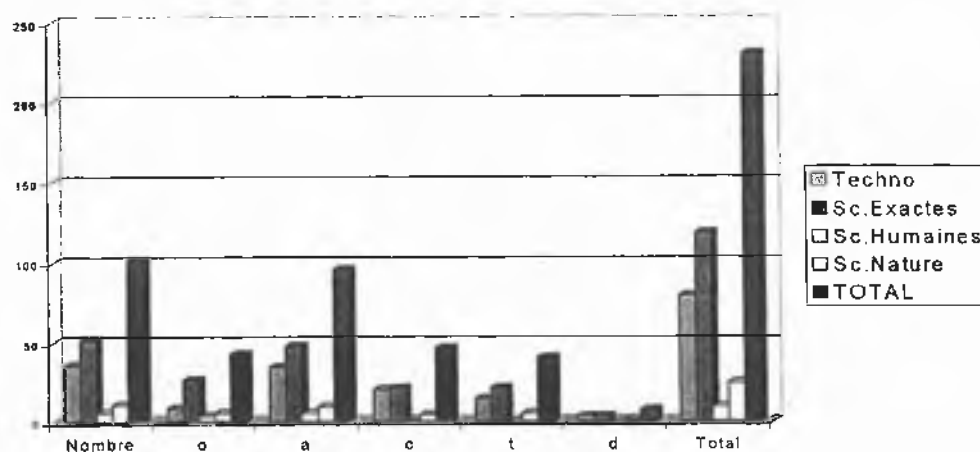
The place of grey literature in documentary research



We notice that the documents of grey literature that are mostly required are the "universitary documents": theses, papers, research reports (especially those related to the university research units).

Moreover, we notice that if the survey has shown that the technological and natural sciences are predominating in terms of the Internet use of grey literature, the human sciences, on the opposite, are practically absent in this field.

The use of Internet and grey literature in the different disciplines



4. COMMENTS ON THE RESULTS OF THE SURVEY

4.1. The appearance of Internet at the university seems to have enhanced an Infatuation for the electronic research among the teachers-researchers. We suppose that there are several reasons:

- a. For many, Internet will solve all the problems related to the research and access to scientific and technical information. This attitude is a little bit naïve when we take into account certain assumptions that qualify this tool as "unique" and "without any disadvantage".
- b. The use of Internet free of charge is another main reason of its exploitation.
- c. For certain persons, it may be a fad to navigate and surf, a manner to show that they move with the movement of fashion.
- d. For the majority of teachers, Internet is a chance to be integrated in the network of electronic communication. This electronic mail is tremendously mentioned and allows the questioners :
 - To meet other colleagues from other countries.
 - To be known and to make their research works known.
 - To publish their works in famous international reviews.

We have been convinced – during this survey- that a majority of teachers know what they look for because they know well this technological tool that they manipulate – in Algeria or abroad – for a relatively long time.

We should mention here that according to the person in charge of the service of the SERIST at the university of Constantine about 1000 teachers are homely connected to Internet.

4.2. This survey has also allowed us to know the place of grey literature in electronic research. It seems that the documents that are mostly required are those that are necessary for research and teaching activities – the article published in periodicals has got a priority – this document will certainly remain a pre-requisite research tool for the world scientific community.

The documents of grey literature, mainly university documents : theses, papers, reports are documents that – like any document of grey literature – show difficulties to have access to.

Does this type of resources satisfy teachers? This should be the object of another survey.

CONCLUSION

It may be too early to express one's opinion about the effective usage of the Internet by the teachers-researchers of the university of Constantine.

It is a recent means whose mastery is not totally acquired. However, through our modest inquiry, the premises are already significant.

Because it favors the research of technical and scientific information in a fast, efficient and free manner, the Internet will constitute, in the future, an incomparable means of scientific communication.

For the moment, the practices of the electronic research seems to be the exclusive fact of teachers stemmed from scientific or technical fields – as well for searching documents of grey literature.

What confirms our hypothesis, according to which there would be a predominance of scientific disciplines and an absolute absence of social sciences.

The most important is to give a following to this work: carry out inquiries to follow the evolution of the Internet usage by the teachers-researchers of Constantine and focusing on important part which we did not mention in our inquiry . The place of the information specialist as a mediator in these new practices of research.

EMERALD INTELLIGENCE + FULLTEXT

WWW.EMERALD-LIBRARY.COM

- ◆ Access to over 100 management journals
- ◆ Over 25,000 articles
- ◆ Liberal network licence

The Management Library on your Desktop

Emerald Intelligence + Fulltext is a current subscription and archived database of over 100 of the world's most important management journals. Respected journals such as the *European Journal of Marketing*, *Management Decision*, *The TQM Magazine*, *Library Management* and the *Personnel Review* are included. That's thousands of fully referenced, serious scholarly articles covering most of what you ever wanted to know and quite a lot you didn't know you wanted to know too.

Emerald Intelligence + Fulltext is available online and on CD-ROM. A liberal network licence is included, so access can be provided for any number of users within an organization. Quick and easy access to articles is assured by Emerald's sophisticated search engine and by the comprehensive labelling of all articles for quality, type and content. The full text of all articles are available instantly in both HTML and PDF formats from 1994 to date. Abstracts of articles are also available from 1989 to 1993 (where available) the full text of which is available from our archive collection.

Emerald Intelligence + Fulltext - applications include:

product/service development research • industry trend analysis •
training /development/learning support • professional, knowledge/expertise
development • customer/competitor analysis • strategic development

Knowledge is power

Welcome to the power station, **Emerald Intelligence + Fulltext**

**For more information or a free 30 day trial please visit
the website, alternatively contact:**

Paul Eddershaw, MCB University Press, 60/62 Toller Lane, Bradford,
West Yorkshire BD8 9BY UK

Tel: +44 (0) 1274 777700
Fax: +44 (0) 1274 785201
E-mail: peddershaw@mcb.co.uk



List of Participating Organizations

AWI	Science and Innovation Administration	BE
BCOHTA	British Columbia Office of Health Technology Assessment	CA
BIOSIS	Publisher of Biological Abstracts and Zoological Record	US
BLDSC	British Library Document Supply Centre	UK
-	Camera dei Deputati, Biblioteca	IT
CASI	Center for AeroSpace Information, see NASA	US
CCC	Copyright Clearance Center	US
CERIST	Centre de Recherche sur l'Information Scientifique et Technique	DZ
CERN	European Centre for Higher Energy Physics, Library	CH
CIAO	Columbia International Affairs Online, Columbia University Press	US
CLRC	Central Laboratory of the Research Councils	UK
CNR	Consiglio Nazionale delle Ricerche	IT
CNRS	Centre National de Recherche Scientifique, see INIST	FR
CORD	Centre on Rural Documentation	IN
-	Cornell University	US
CPS	Centre for Property Studies	CA
-	CYBEK of New York	US
DOE	U.S. Department of Energy, see OSTI	US
DTIC	Defense Technical Information Center	US
EAGLE	European Association for Grey Literature Exploitation	UK
EMICH	Eastern Michigan University	US
FID	International Federation for Information and Documentation	CA
FIT	Florida Institute of Technology	US
GREYNET	Grey Literature Network Service	NL
GSFC	Goddard Space Flight Center, See NASA	US
GUKCC	Gallaudet University Kellogg Conference Center	US
-	Harvard University	US
HP	Hewlett Packard Company	US
HUT	Helsinki University of Technology	FI
ICSSR	Indian Council of Social Science Research	IN
IDA	Institute for Defense Analyses	US
IFLA	International Federation of Library Associations & Documentation	UK
-	Informed Strategies	US
INIST	Institut de l'Information Scientifique et Technique	FR
IRC	Information Resource Center	US

Anbar International Management Database

Anbar International Management Database is undoubtedly the world's best management library today ! For more than forty years it has provided a fast track to the effective application of knowledge.

The database contains an abstract for every article published in more than 400 of the world's most influential journals, building on an archive stretching back to 1989. That means access to over 90,000 management abstracts. The full text is available immediately online for almost a quarter of the accredited journals, the remainder can be ordered using our document delivery service (DDS), via our partner, the British Library.

The accreditation board which includes leading figures, such as Philip Kotler of Northwestern University and Alfred Lynch, CEO, J C Penny International Inc, ensure that Anbar's coverage does not miss any significant thinking and practice. In reaching their decision they examine the library holdings of centres of excellence including Harvard University and MIT

◆ Areas of Coverage

The entire management field is comprehensively covered by the *Anbar International Management Database* to ensure that you do not miss significant new thinking and practice,

Coverage includes:

- Accounting & Finance
- Information Management & Technology
- Management of Quality
- Marketing & Logistics
- Operations & Production Management
- Human Resource Management

◆ What the Service Delivers

Subscribers to the *Anbar International Management Database* receive a comprehensive package, including:

- Current and archival access to the entire database of over 400 journals via the Internet
- CD-ROMs of archived material updated three times a year.
- Liberal network licence providing access for any number of users within the organization
- Search engine using article type, keyword, author and more, allowing rapid retrieval of articles
- Bibliographic information for a further 1200+ journals taken from the International Directory of Periodical Contents in Management
- Six hard copy current awareness journals packed with articles reviews
- Full text retrieval service (within 2 working hours by fax or 24 hours by post) of any article in the database
- Three issues of the Anbar magazine *Intelligence*

◆ What Next ?

Don't take our word for it ! For a FREE 30 Day Trial log on to

www.anbar.com

If you require any additional information on our range of Anbar products please contact:

Paul Eddershaw

Telephone: +44 (0) 1274 777700

Facsimile: +44 (0) 1274 785201

E-mail: anbar@anbar.co.uk


ELECTRONIC INTELLIGENCE

Author Index

A-B

✓ Aceti, A.	61
✓ Aina, L.O.	25
✓ Almeida, M.	181
✓ Artus, H.M.	12
✓ Blaaij, C.A.T. de	143

M

✓ Maeda, Y.	282
✓ Maly, K.	78
✓ McCampbell, A.S.	203
✓ McMillan, G.	32
✓ Mili, F.	238
✓ Morita, I.	170
✓ Musser, L.R.	227

C

✓ Castriotta, M.	61
✓ Cho, H-Y.	217
✓ Choi, S-H.	217
✓ Choi, S-P.	217
✓ Clare, L.M.	203
✓ Cornish, G.P.	164
✓ Cutler, D.	109

N - O - P - R

✓ Nakagawa, K.	282
✓ Nelson, M.L.	78
✓ Osif, B.A.	47
✓ Rama Devi, T.	260

D - G

✓ Davis, D.	156
✓ Dallman, D.	230
✓ Di Cesare, R.	61, 297
✓ Farace, D.J.	iii
✓ Gelfand, J.	1
✓ Gitters, S.H.	203
✓ Green, C.	276

S

✓ Sato, H.	282
✓ Savoie, I.	276
✓ Schopf, J.	199
✓ Seadle, M.	135
✓ Semra, H.	307
✓ Shim, H-S.	217
✓ Siwek, K.	254

H - J - K - L

✓ Helmer, D.	276
✓ Jeffery, K.G.	88
✓ Lazzari, G.	297
✓ Le Meur, J-Y.	230
✓ Lodi, E.	116
✓ Luzi, D.	61

T-V

✓ Takalkar, A.	260
✓ Vesely, M.	116
✓ Vigen, J.	116

Publication Order Form

INFORMATION RESOURCES

Nº OF COPIES X AMOUNT IN EUROS

GL'99 Conference Proceedings	ISBN 90-74854-24-9	_____ x € 55.00 = _____
GL'99 Program & Abstracts	ISBN 90-74854-22-2	_____ x € 15.00 = _____
GL'97 Conference Proceedings	ISBN 90-74854-17-6	_____ x € 55.00 = _____
GL'97 Program & Abstracts	ISBN 90-74854-18-4	_____ x € 15.00 = _____
GL'95 Conference Proceedings	ISBN 90-74854-08-7	_____ x € 55.00 = _____
GL'95 Program & Abstracts	ISBN 90-74854-09-5	_____ x € 15.00 = _____
GL'93 Conference Proceedings	ISBN 90-74854-03-6	_____ x € 55.00 = _____
GL'93 Program & Abstracts	ISBN 90-74854-04-4	_____ x € 15.00 = _____
GLOSSARY, a vocabulary of terms in GL	ISBN 90-74854-19-2	_____ x € 35.00 = _____
Annotated Bibliography on Grey Literature	ISBN 90-74854-20-6	_____ x € 30.00 = _____
Int. Guide to Persons and Organisations in GL	ISBN 90-74854-23-0	_____ x € 30.00 = _____
Notebook on Grey Literature	ISBN 90-74854-15-X	_____ x € 35.00 = _____
NewsBriefNews: Quarterly Newsletter on GL	ISSN 1389-1804	_____ x € 20.00 = _____

Postage & Handling (Holland) € 5.00 = _____

Postage & Handling (Europe) € 10.00 = _____

Postage & Handling (Other) € 15.00 = _____

SUBTOTAL = € _____

☐ Reduction of 15 % for GreyNet Members

= € _____

TOTAL = € _____

Name: _____

Organisation: _____

Address / P.O. Box: _____

P-code/City/Country: _____

Please Check one of the boxes below for the method of payment

☐ Bank check/draft enclosed and payable to GreyNet.

☐ Please remit to GreyNet | ABN♦AMRO Bank No. 52.93.61.841 in Amsterdam, Netherlands.

☐ MasterCard/Eurocard ☐ Visa/JCB Card ☐ American Express ☐ Diners Club Card

No. _____ Expiration Date: _____

If the name on the credit card is not the same as mentioned above, print the name that appears on the card here _____

Signature: _____

Date: _____

Please Mail or Fax this Signed Form to: GreyNet, Grey Literature Network Service,
Koninginneweg 201, 1075 CR Amsterdam, The Netherlands
TEL/FAX: 31-20-671.1818 - Email: GreyNet@inter.nl.net