

WORK ON GREY IN PROGRESS

SIXTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

HOST AND SPONSORS

New York Academy of Medicine
United States

NYAM

Institute for Scientific and Technical Information
France

INIST-
CNRS

Endeavor Information Systems Inc.
United States

Endeavor

PROGRAM COMMITTEE

New York Academy of Medicine
United States

NYAM

Research Libraries Group
United States

RLG

University of Bergen
Norway

UiB

Grey Literature Network Service
Netherlands

GreyNet

University of California, Irvine
United States

UCI

National Research Council
Italy

CNR

University of Twente
Netherlands

UTwente

Institute for Scientific and Technical Information
France

INIST

Japan Science and Technology Agency
Japan

JST

Columbia University
United States

Columbia

NEW YORK ACADEMY OF MEDICINE
NYAM CONFERENCE CENTER
NEW YORK, N.Y., USA

6-7 DECEMBER 2004



Conference Proceedings

ISSN 1386-2316

TextRelease

GL6 Program and Conference Bureau
Beysterveld 251, 1083 KE Amsterdam, NL

Tel/Fax +31-(0) 20-672.1217

info@textrelease.com
www.textrelease.com

CIP

GL6 Conference Proceedings

Sixth International Conference on Grey Literature : Work on Grey in Progress, 6-7 December 2004 / compiled by D.J. Farace and J. Frantzen. - Amsterdam : TextRelease, December 2004. - 182 p. ; 30 cm. - Author Index. - (GL-conference series, ISSN 1386-2316 ; No. 6)

The conference proceedings contain the full text of some twenty-five papers presented during the Conference Sessions and Panels. The print copy has been standardized in font, size, and format for general layout and appearance. Included is an Author index and full address information.

ISBN 90-77484-04-3 : € 70

© TextRelease, 2005

Foreword

Work on Grey in Progress

In the transition to the 21st century, grey literature has become a daily work experience in public, academic, and business sectors. In the span of a decade - owing to Internet technology - grey literature has moved from uncharted and untapped resources to a mainstream source of information. This information revolution has tremendous impact on the landscape and environment in which grey literature is produced, processed, and exploited. Workflow patterns embedded in grey require innovation by new and longstanding stakeholders.

These proceedings provide the full-text of some twenty-four papers presented at the Sixth International Conference on Grey Literature, which was hosted by the New York Academy of Medicine.

GL6 offered information professionals a global platform on which to showcase best practices and benchmarks i.e. "work on grey in progress". Delegates and conference participants from thirteen countries worldwide and thirteen states nationwide attended this annual event, where more than sixty-five organizations were represented.

On behalf of the authors, I invite you to read these papers and genuinely welcome your thoughts and ideas.

Dr. Dominic J. Farace
Grey Literature Network Service

Amsterdam,
January 2005

GL6 Conference Sponsors

NYAM

New York Academy of Medicine, USA



INIST-CNRS

Institut de l'Information Scientifique et Technique;
Centre National de Recherche Scientifique, France



Endeavor

Endeavor Information Systems Inc., USA



GL6 Program Committee Members and Organizations



Janice Kaplan

New York Academy of Medicine, USA



Constance Malpas

Research Libraries Group, USA



Anne Asserson

University of Bergen, Norway



Dr. Dominic J. Farace

Grey Literature Network Service, Netherlands



Julia Gelfand

University of California, Irvine, USA



Daniela Luzzi

National Research Council, Italy



Prof. Dr. Hans E. Roosendaal

University of Twente, Netherlands



Dr. Joachim Schöpfel

Institut de l'Information Scientifique et Technique, France



Hideki Takahashi

Japan Science and Technology Agency, Tokyo, Japan



Kate Wittenberg

Columbia University, USA

Contents

Foreword	iii
Sponsors and Program Committee Members	iv
Conference Program	vi-vii
Opening Session <i>Grey Literature Survey 2004</i>	1-9
Session One <i>Institutional Repositories</i>	10-34
Session Two <i>Use Analysis</i>	35-63
Session Three <i>IT & Research Initiatives</i>	64-83
Panel One <i>Knowledge Management and Dissemination</i>	84-122
Panel Two <i>Collection Development and Resource Discovery</i>	123-176
Appendices	
List of Participating Organizations	178-179
GreyNet Subscription Form	180
Index to Authors	181
GL6 Publication Order Form	182

Opening Session

Chair: J. Kaplan, New York Academy of Medicine, NYAM – United States

- A.K. Boekhorst, University of Amsterdam – Netherlands; University of Pretoria** 1
D.J. Farace, Grey Literature Network Service and J. Frantzen, Boer & Croon – Netherlands
 Grey Literature Survey 2004: A Research Project Tracking Developments in the Field of Grey Literature

Session One – Institutional Repositories

Chair: A. Asserson, University of Bergen, UiB - Norway

- J. Gelfand, University of California, Irvine – United States** 10
 "Knock, Knock:" Are Institutional Repositories a Home for Grey Literature?
- E.M.S. Dijk, NIWI/KNAW – Netherlands** 16
 Sharing Grey Literature by using OA-x
- A.C. Japzon and N. Anderson, IIA, GSFC/NASA – United States** 24
 Wallops Island Balloon Technology: Can't see the repository for the documents
- N. Rupp and L.J. LaFleur, Cornell University – United States** 30
 Making Grey Literature Available through Institutional Repositories

Session Two – Use Analysis

Chair: D. Luzi, National Research Council, CNR/IRPPS - Italy

- J.M. Barrueco, University of Valencia – Spain** 35
T. Krichel, Long Island University – United States
 Building an autonomous citation index for grey literature: the Economics working papers case
- R. Di Cesare, R. Ruggieri, S. Giannini, and S. Biagioni, CNR – Italy** 41
 Trend evaluation and comparison of the use and value of GL in core demography and computer science journals
- K. Dunn, Ehrman Medical Library; NYU, United States** 50
 Impact of the Inclusion of Grey Literature on the Scholarly Communication Patterns of an Interdisciplinary Specialty
- J. Schöpfel and C. Stock, INIST-CNRS, France;** 55
D.J. Farace, GreyNet; J. Frantzen, Boer & Croon, Netherlands
 Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit

Session Three – IT & Research Initiatives

Chair: J. Schöpfel, Institut de l'Information Scientifique et Technique, INIST/CNRS - France

- D.E. Cutler, OSTI/DOE – United States** 64
 Grey Literature in Energy: Five Years Later
- C. Bourasseau and C. Dumas, Ecole des Mines des Nantes – France** 68
 A study of teachers and researchers practices with digital documents – grey or not
- M. Cooke and S. Costigan, Columbia University Press - United States** 79
 Making Your Way Through Grey: Metadata, MARC and User Tools

Panel One – Knowledge Management and Dissemination: Part I

Chair: K. Wittenberg, Columbia University; Electronic Publishing Initiative at Columbia – United States

- K.G. Jeffery, CCLRC – Rutherford Appleton Laboratory - United Kingdom** 84
A. Asserson, University of Bergen, UiB - Norway
 Relating Intellectual Property Products to the Corporate Context
- P. Simkhada, J. Knight, G. Reid, L. Wyness and L. Mandava, Aberdeen Medical School – UK** 91
 Chasing the grey evidence: A standardised Systematic Critical Literature Review Approach
- D. Luzi, CNR/IRPPS, M. Castriotta, ISPESL and M. Manco, and R. Mazzucco, LINK – Italy** 94
 The integration of GL documents with the research information system on occupation safety and health

Panel One – Knowledge Management and Dissemination: Part II

Chair: K. Wittenberg, Columbia University, Electronic Publishing Initiative at Columbia – United States

- | | |
|--|------------|
| C. de Blaaij, Library of Zeeland, Academic Department – Netherlands | 106 |
| Two worlds - about bars and stars in scientific information publishing; An analysis of open source ideology as a means of self-controlled publishing | |
| L.P. Pavlov, Scientific and Technical Information Centre of Russia, VNTIC – Russia | 113 |
| The problems of grey in the context of postmodern science and culture | |
| H. von Hofe, CYBEK – United States | 119 |
| Towards a Genealogy of Grey Literature via Newton's Journals | |

Panel Two – Collection Development and Resource Discovery: Part I

Chair: J. Gelfand, University of California, Irvine – United States

- | | |
|---|------------|
| L. Myohanen, E. Taylor and L. Keith - New York Academy of Medicine, NYAM – USA | 123 |
| Accessing Grey Literature in Public Health: New York Academy of Medicine's Grey Literature Report | |
| S.L. Ranger, University of Houston, UH – United States | 128 |
| Grey Literature in Special Libraries: Access and Use | |
| H. Lehman, University of Washington Information School – United States | 136 |
| J. Webster, Oregon State University Libraries – United States | |
| Describing grey literature again: A survey of collection policies | |

Panel Two – Collection Development and Resource Discovery: Part II

Chair: J. Gelfand, University of California, Irvine – United States

- | | |
|--|------------|
| P. Paillassard, J. Schöpfel and C. Stock, INIST-CNRS – France | 143 |
| How to get a French doctoral thesis, especially when you aren't French | |
| J. Carman, A. Creel, S. Milewski, J. Staub and A. Woofter | |
| The School of Information Sciences; University of Tennessee, Knoxville, UTK – USA | 160 |
| ORNL-Safeguards and security: The Global producers of the 8 radionuclides of particular security concern | |
| R.M. Jackson, Portland State University, College of Urban and Public Affairs – USA | 169 |
| Grey Literature and Urban Planning: History and Accessibility | |

Connecting to your world of resources

Endeavor Information Systems

delivers pioneering, comprehensive library management and information technologies to leading academic and research libraries. Nearly 1,300 Endeavor customers around the world help users connect to information every day.

- **Voyager™** integrated library management system
- **ENCompass™**, the digital library system for organizing and searching collections
- **LinkFinderPlus™** OpenURL-enabled link resolver
- **Endeavor Meridian™** system for efficient, effective electronic resource management

 **Endeavor**
Information Systems Incorporated

www.endinfosys.com • +1 (847) 296-2200

United States & Canada
(Headquarters)
info@endinfosys.com

Latin America
info.sa@endinfosys.com

United Kingdom
info.uk@endinfosys.com

Europe, Middle East
& Africa
info.europe@endinfosys.com

Asia
info.asia@endinfosys.com

Australia & New Zealand
info.austnz@endinfosys.com

Grey Literature Survey 2004

A research project tracking developments in the field of grey literature

A.K. Boekhorst, University of Amsterdam, UvA; University of Pretoria, UP

D.J. Farace, GreyNet, Netherlands

J. Frantzen, Boer & Croon, Netherlands

Introduction

Last December, at GL5 in Amsterdam, many of the authors and researchers reiterated the Luxembourg Convention on grey literature. Some questioned if it were not time to rethink the definition, some offered moderations, but not one called for its complete elimination or abolition. During the course of the authors' presentations, specific attributes of grey were brought to lively discussion and debate. However, time and circumstance being known limitations on this kind of interaction and communication, ensured that more issues and questions were raised than could be adequately addressed. At the close of GL5, the idea came to devise a research project capturing the main issues and questions raised and to systematically hold them up to the Luxembourg definition of grey literature to examine where redefinition if any is needed. Both authors are well aware of crossovers, bypasses, and at times a *Fata Morgana* as to what is grey or commercial on the information highway, but then they poise the question, where would be today without definition?

The dichotomy grey versus commercial is one found both in the definition following the York Seminar (Wood, 1984) "...not available through normal bookselling channels" and the definition redefined at the Luxembourg Conference (1997) "... not controlled by commercial publishing". Like any dichotomy, this one serves as a cognitive tool in the process of reduction, analysis, and understanding of information - from authorship and production to its further use and application. This dichotomy also allows for the simplification of widespread terminologies, where grey literature becomes an umbrella term for a host of expressions among which include the terms: secondary, unpublished, unconventional, fugitive, refugee, fringe, minor, open, and etcetera.

Further, this dichotomy has allowed information workers to group the types of documents such as reports, working papers, proceedings, and 100s of other document types that are grey literature; and, in so doing set them apart from what is commercial literature. In this same process of typology, grey literature distinguishes itself from yet another category of information, which includes ephemera, black literature, vanity press, web trash, classified information, etc. Thus, the dichotomy allows one to distinguish not only between that which is grey and commercial but also it excludes that which belongs to neither.

Like any tool or instrument, its purpose serves as a means to an end, and is not the end itself. The dichotomy grey versus commercial has allowed information professionals to continue their work in progress without becoming lost in the mass and gigabytes of information available in every conceived means of print and electronic format. Furthermore, if this definition continues to have meaning for information professionals, then should the average net-user also at least recognize the term grey literature?

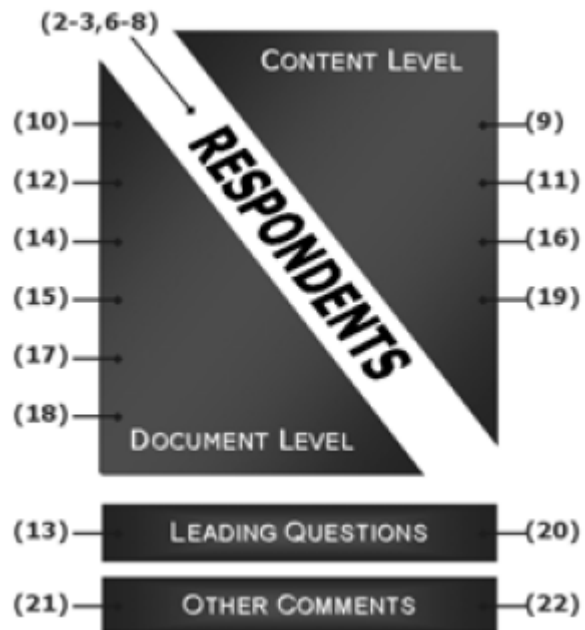
Method and Procedure

This survey was held in advance of the Sixth International Conference on Grey Literature (GL6). The instrument used is an open-ended questionnaire consisting of twenty-two questions or items designed to compile information for further analysis with the aid of SPSS software. It is not only important to know if a respondent agrees or disagrees with the questions, but also their comments stand central. While this questionnaire is not anonymous, the identity of an individual respondent will remain undisclosed. An estimated 10-12 minutes of the respondent's time is needed to complete the online questionnaire.

The questionnaire was mounted to a Webpage on GreyNet's website in late February 2004, where it remained online for 7 months - yielding 104 completed and valid forms. As the forms arrived via email, they were recorded by the project administrator and sent in batches of roughly 20 each to the data controller, where they were further entered and processed via SPSS (Statistical Package for the Social Sciences) software. The first online questionnaire was completed on February 25, 2004 and the last was completed on September 20, 2004. During this period, the project workers met on five occasions to examine preliminary results and address the responses in the open-ended questionnaire that needed to be further grouped and labelled. By the end of October, the final results of the questionnaire had been tabulated and were ready for further analysis and interpretation.

Questionnaire

1. Your name and email address?
2. Country of residence?
3. Sector in which you are currently employed (government, academics, business, industry, or other)?
4. URL of your business or organisation?
5. Name of your department?
6. Title or position you hold within the department?
7. How many years ago did you first encounter the term 'grey literature'?
8. Way(s) you are involved in grey literature (production, processing, distribution, and/or other)?
9. Have you authored one or more publications (article, paper, report, etc.) on the topic of grey literature?
10. Does your business/organisation have a policy or position statement on grey literature?
11. Grey literature is always subject to a review process?
12. Grey literature is best described by the type of document it embodies (e.g. thesis, newsletter, report, website, database, etc.)?
13. The average net-user should at least recognize the term 'grey literature'?
14. Grey Literature collections are better managed by institutional (centralized) than disciplinary (decentralized) repositories?
15. Once grey literature is bibliographically controlled by means of indexing and referencing, it ceases to be grey?
16. The content of commercially published literature is superior to grey literature?
17. Grey Literature should be free to access?
18. Grey Literature should be free of charge?
19. Grey Literature itself constitutes a field in information studies?
20. The Luxembourg Convention on Grey literature (1997) reads, "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing". Does this still hold?
21. In what way could GreyNet better serve the grey literature community?
22. Other comments?

Grey Literature Survey 2004**Data Analysis**

In examining the response from the questionnaire and for purposes of presenting the results of this survey, the questions or items have been grouped and further paired. The 22 survey items now fall under three main headings: profile of the respondents, response to survey items on a document level, and response to survey items on a content level. The response to the leadings questions of definition and exposure to the term grey literature to net-users are then considered. The last two items on the questionnaire are placed under the general heading 'other comments' and can be observed in the final research results. The explorative character of the survey lent itself to open questions, since this offers the possibility of subtle distinctions in the answers. The respondents extensively made use of this. Many answers to the questions were not given in the form of a clear 'yes' or 'no', but had a narrative character. In order to facilitate the statistical analysis, the data were categorized in groups at several steps during the course of the research project. And, in a final step, the answers were recoded into: 'Yes', 'No', 'Depends', 'No Answer', and 'Rhetorical Answer'. Ultimately, for the final analysis 'No Answer' and 'Rhetorical Answer' were combined under the heading 'NA'.

Profile of the Respondents

If we look at when the respondents first encountered the term 'grey literature' i.e. the length of time exposed to grey literature and the spread of the term over the years, a curve presents itself. On either end are fewer than 10% of the respondents and in each of the four periods in between there is an average of 20% of the response. The mean response lies between 6-10 years and a span of more than a

quarter of a century exists among the respondents. Looking at this curve we might ask ourselves if 'grey' bridges the proverbial information gap? And, if so, to what extent is the Internet responsible for this?

Variable 7: How many years ago did you first encounter the term 'grey literature'?

	<i>Frequency</i>	<i>Percent</i>
1 yr or less	6	5,8
2-5	22	21,2
6-10	25	24,0
11-15	20	19,2
16-20	17	16,3
21 and more	10	9,6
NA	4	3,8
Total	104	100,0

When we view the respondents in relation to their first contact with GreyNet (Grey Literature Network Service) on whose website the questionnaire was mounted, we find a somewhat similar curve, where less than 20% of the respondents are on either end and the two categories in the middle account for equal numbers of respondents, just under 35% each.

Comparative data: Relationship of the Survey Respondents to GreyNet

17.3% had contacts with former Greynet <2000
34.6% had first contact with Greynet after its re-launch in 2003
34.6% had first contact with GreyNet via the Survey 2004
13.5% had first contact with GreyNet in 2004 and afterwards completed the Survey
100%

The 104 survey respondents were from twenty-six countries worldwide. However, those from Europe and North America accounted for more than 85% of the response. If we examine the countries of residence further, we find that 75% of the respondents came from six of the twenty-six countries and that two of those six countries accounted for a little over 50% of the total response, namely the USA (34.6%) and the Netherlands (19.2%). This could perhaps be explained by the fact that the Survey was in lieu of a conference that would be held in New York and the home office for the survey was in Amsterdam. In any event, the concern is not with the over representation of Europe and North America or with The States and The Netherlands, but rather with the under representation of Africa, Asia, and South America, which together accounted for less than 15% of the survey response.

Variable 2: Country of residence?

	<i>Frequency</i>	<i>Percent</i>
North America	39	37,5
Europa	50	48,1
Other	15	14,4
Total	104	100,0

When we examine the sectors in which the respondents are currently employed, we find that Academics accounts for a little over 50% of the response with Government at about 30%. NGOs, IGOs, and Foundations, which together form the 'Other' category, amounted to 11.5% of the response, which is still twice as much as the response from the Business sector that ended with less than 5%. If we compare this with GreyNet statistics from six years ago, where 26% of its contacts were with Business and Industry (NewsBriefNews, vol. 7, no.3), than an alarming decrease among these stakeholders appears to have occurred. The concern here is then not with an over representation of Academics, but with a severe under representation of business and industry. Have the firewalls in networked environments become the new obstacles to acquiring and accessing business grey? And, on the other hand, has academics found its true home in grey?

Variable 3: Sector in which you are currently employed (government, academics, business, industry, or other)?

	Frequency	Percent
Academic	56	53,8
Government	31	29,8
Business	5	4,8
Other	12	11,5
Total	104	100,0

If we look at the titles or positions that the respondents hold in their respective departments, we find that just fewer than 30% are involved in technical and/or operational capacities, while nearly 35% are working on managerial levels, and almost 34% are in teaching and research.

Variable 6: Title or position you hold within the department?

	Frequency	Percent
Managerial	36	34,6
Technical/Operational	31	29,8
Teacher/Researcher	35	33,7
NA	2	1,9
Total	104	100,0

When we turn to the way the respondents are involved in grey literature, taking into consideration that their involvement can be in more than one way, we see that processing and distribution of grey literature, which are the traditional activities out of which grey literature arose, occupy roughly 60% of the respondents, while near 40% of the respondents' occupation with grey literature is in production, management, teaching, and research. Furthermore, an average of 60% of the survey respondents had multiple affinities with grey literature.

Variable 8: Way(s) you are involved in grey literature (production, processing, distribution, and/or other)?

	Count	Pct of Responses	Pct of Cases
Production	33	19,6	32,0
Processing	63	37,5	61,2
Distribution	39	23,2	37,9
Other	33	19,6	32,0
Total responses	168	100	163,1

If we compare this with the results of a survey held by GreyNet a decade ago in which 142 respondents participated (*NewsBriefNews*, vol. 3, no.2), we find that 70% had a multiple affinity with grey literature and that roughly the same percentage were involved in traditional activities with grey literature, namely processing and distribution. Granted that it is very difficult to make comparisons with data gathered by two different means with a ten-year interval, still however, the response of the GL Survey 2004, lead us to believe that more information professionals are occupied in the management, research, and teaching of grey literature than a decade ago. And that there is some increased levels of specialization. Multiple involvements with grey literature appear to remain at the same level; only the type of involvement has shifted.

On the Document Level

If we look at the response to the survey items on a document level, we find that two-thirds of the respondents agree that grey literature is best described by the type of document it embodies. In other words, a type of document does not define grey literature, but it does help to identify grey literature. For example, preprints, working papers, reports, etc. are in almost all cases identified as grey literature i.e. not controlled by commercial publishing.

Variable 12: Grey literature is best described by the type of document it embodies (e.g. thesis, newsletter, report, website, database, etc.)?

	Frequency	Percent
Yes	69	66,3
Depends	6	5,8
No	22	21,2
NA	7	6,7
Total	104	100,0

When we compare the response to the above with the response to whether a document once bibliographically controlled ceases to be grey, we find a near 70%, who disagree. For these respondents, bibliographic control is not the determining factor whether a document is grey or not, which seems to support the notion that grey literature is first and foremost determined by the publishing or issuing body – be it a grey or commercial publisher. Roughly 15% agreed and the other 15% either chose not to respond or their response was such that it could not be taken as a simple yes or no.

Variable 15: Once grey literature is bibliographically controlled by means of indexing and referencing, it ceases to be grey?

	Frequency	Percent
Depends	9	8,7
No	70	67,3
Yes	16	15,4
NA	9	8,7
	104	100,0

If we turn to the question as to whether grey literature should be free to access, we once again see a solid response of 60%, who agree. While another 20% agree to some extent and less than 15% who disagree. Today, with the OAI, Open Archive Initiative, it is not unlikely that the majority of respondents would be in favor of this.

Variable 17: Grey Literature should be free to access?

	Frequency	Percent
Yes	64	61,5
Depends	21	20,2
No	15	14,4
NA	4	3,8
Total	104	100,0

When, we compare the results of the above with the question as to whether grey literature should be free of charge, we again see a majority of near 50% who agree and a near 20% who disagree, while a near 25% hinge their response on further clarification. It is the response of this quarter of the respondents that must be further examined. While grey literature by current definition is not commercially attractive to publishers, there are still costs involved in its production, processing, and distribution, which must be recovered. Perhaps the disparity can be further explained by the sector in which the grey literature originates. Government and Academics in all likelihood have a different policy than non-governmental agencies, and certainly business and industry. Since nearly 85% of the respondents to this survey come from government and academics, then this may explain the majority of response to the question.

Variable 18: Grey Literature should be free of charge?

	Frequency	Percent
Depends	25	24,0
No	19	18,3
Yes	51	49,0
NA	9	8,7
Total	104	100,0

If we look at the results as to whether or not a policy exists in ones organization, which addresses grey literature, then a near 70% do not have such a policy, while more than 25% do. This survey question was sparked by a recent study carried out at Portland State University (Siegel, 2004). Organizations producing, processing, and/or distributing grey literature need to formalize their policy towards their own grey literature and it should be recorded and made public.

Variable 10: Does your business/organisation have a policy or position statement on Grey literature?

	<i>Frequency</i>	<i>Percent</i>
Yes	28	26,9
No	72	69,2
NA	4	3,8
	104	100,0

When we examine the response as to the best home for grey literature collections, we see a relative split on whether institutional or disciplinary repositories are more suited. Both have more than 30% of the response and another near quarter of the respondents is less emphatic in their response by elaborating on the pros and cons of each. Another +10% chose not to respond. This issue is certainly not settled and while institutional repositories are in the current limelight, advocates of decentralized and disciplinary repositories remain outspoken.

Variable 14: Grey Literature collections are better managed by institutional (centralized) than disciplinary (decentralized) repositories?

	<i>Frequency</i>	<i>Percent</i>
Depends	25	24,0
No	33	31,7
Yes	35	33,7
NA	11	10,6
Total	104	100,0

On the Content Level

Looking at the content level of grey, we find that almost 45% of the respondents agree that grey literature is always subject to a review process, while less than 40% disagree and a near 10% are caught in the middle and are less prepared to answer with a simple yes or no. In much the same way as an institution should formulate a policy on grey literature, so also should the author or corporate author be steadfast in identifying to the reader the review process under which their grey publications have passed. For some types of grey literature this may already be implicit e.g. a thesis has an academic committee, a grey journal has an editorial and advisory board, a conference paper has a program committee, etc. However, for other types of grey both in print and electronic formats, the review process is not transparent and should be stated explicitly.

Variable 11: Grey literature is always subject to a review process?

	<i>Frequency</i>	<i>Percent</i>
No	41	39,4
Depends	10	9,6
Yes	46	44,2
NA	7	6,7
Total	104	100,0

When we compare to the response to the question as to whether commercially published literature is superior to grey, we see that less than 5% agree, while almost 55% disagree and a solid 30% choose to explain their answers further. It becomes clear that it is not the content, but the packaging, promotion, marketing, and other such value added services, which make commercial publications more attractive. And, in so doing making them more commercially viable. Following this further, it is not unthinkable that an author who first published a paper in a conference proceedings, which later crossed over to a journal article, would choose to cite the journal article instead of the original publication.

Variable 16: The content of commercially published literature is superior to grey literature?

	Frequency	Percent
Yes	5	4,8
Depends	34	32,7
No	57	54,8
NA	8	7,7
Total	104	100,0

In response to whether a respondent had authored one or more publications on the topic of grey literature, we find an even split in the response. Almost half of the survey respondents are in some way meta-authors of grey i.e. they have dealt with one or more aspects of grey literature in research and publication, and have recorded it for later use and application. If we view this in relation to the present GL6 Conference, we find that 26% of the survey respondents are also authors or co-authors. And, if we include the previous conferences in the GL-Series, then this increases to 31%.

Variable 9: Have you authored one or more publications (article, paper, report, etc.) on the topic of grey literature?

	Frequency	Percent
No	52	50,0
Yes	50	48,1
NA	2	1,9
Total	104	100,0

While just a little over 15% of the survey respondents do not agree that grey literature constitutes a field in information studies, a near 70% of the respondents do, and almost 8% who were somewhat unsure clarified their position and appear to lean in favor of grey literature as a field of information studies. When we compare this three-quarter majority of response with the response to survey item 10 (i.e. the Title/Position of the Respondent, where 30% of the respondents are in teaching and research) then it is not unlikely to assume that their activities are already related to grey literature as a field of information.

Variable 19: Grey Literature itself constitutes a field in information studies?

	Frequency	Percent
No	16	15,4
Perhaps	8	7,7
Yes	72	69,2
NA	8	7,7
Total	104	100,0

Research Results

When we finally arrive at the results from the two leading questions in the survey - one of definition and the other of terminology - we are not be surprised by the response. More than 70% of the survey respondents maintain the Luxembourg definition on grey literature and another 5% while holding to the definition offer some modification. Less than 15% of the respondents do not maintain the definition. This leading question rightfully had its place at the end of the questionnaire, i.e., once the respondent had already addressed a wide range of aspects and notions about grey literature. If there were doubt, then one might fairly assume that this leading question would not have scored the highest percentage of response (72.1%) than any of the other survey questions received.

Variable 20: The Luxembourg Convention on Grey literature (1997) reads, "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing". Does this still hold?

	Frequency	Percent
No	14	13,5
Yes	75	72,1
Yes, but need revision	5	4,8
NA	10	9,6
Total	104	100,0

When we examine the response to the second and final leading question as to whether the average net-user should recognize the term grey literature, exactly 50% thought that they should, while around 35% did not. It appears inevitable that the term will become widespread in use based on the premise that grey literature is now evolving into a field of information. And eventually, it will make its way from keyword lists and specialized glossaries to the Oxford and Webster Dictionaries somewhere between 'grey area' and 'grey matter'.

Variable 13: The average net-user should at least recognize the term 'grey literature'?

	<i>Frequency</i>	<i>Percent</i>
No	37	35,6
Depends	7	6,7
Yes	52	50,0
NA	8	7,7
Total	104	100,0

The last two items on the questionnaire, which were outside of the mainstream of the survey yielded considerably less response than the first twenty items on the questionnaire. To question 21: In what way could GreyNet better serve the grey literature community? Only 87 of the 104 respondents (83.7%) made comments. Some of the respondents maintained that GreyNet, the Grey Literature Network Service, should

- Address its role in the Open Access Movement
- Assist in the best practices tailored to different information communities
- Increase public awareness of its presence and value
- Re-launch a Journal on Grey Literature
- Etcetera

And, finally to the catchall question 22: Other comments? Only 29 of the 104 respondents (27,9%) commented. Among the varied comments included statements such as:

- A Survey in advance of a conference is a great idea
- Studies on Grey Literature should receive strong research backing
- I found this online survey via Google
- Etcetera

Conclusions and Further Recommendations

In 1993-1994, GreyNet carried out its first survey on grey literature. However, within the past decade, due to Internet technology, grey literature has become a daily work experience in all sectors of society and constitutes a mainstream source of information. The main conclusions from the Grey Literature Survey 2004 indicate that the Luxembourg definition on Grey Literature should be maintained with or without modification, that grey publishers should in some way acknowledge the review process in which their print/electronic documents have undergone, and that colleges and schools of library and information studies should continue to develop curricula programs, modules, and courses on grey literature.

Further recommendations based on the findings indicate that ways should be found to correct the disproportion between geographical and sectoral involvement in Grey Literature, that organizations involved in this field of information should formulate and publicize a policy or position statement on grey literature, and that the raw survey data (i.e. without names, email addresses, and URLs) should be made available to researchers for secondary analysis.

Author Information

Prof.dr. Albert K. Boekhorst
University of Amsterdam/University of Pretoria
Faculty of Humanities, Media Studies
Archives and Information Studies
Turfdraagsterpad 9
1012 XT Amsterdam
The Netherlands

a.k.boekhorst@uva.nl
www.hum.uva.nl/akb/

References

1. Babbie, Earl R. (1975), *The Practice of Social Research*. – Belmont, California : Wadsworth Publishing Co., 1975. – XIV, 511 p. : 24 cm. – ISBN 0-534-00381-8
2. Wood, D.N. (1984), The collection, bibliographic control and accessibility of grey literature. – In: *IFLA Journal*, vol. 10, no. 3, 1984, pp. 278-282.
3. Farace, D.J. ed. (1994), *SPSS Final Results*. – In: *NewsBriefNews*, vol. 3, no. 2, pp. 1-2. – Grey Literature Network Service, GreyNet/TransAtlantic : Amsterdam, 1994. – ISSN 0929-0923
4. Heer, de I. and D.J. Farace (1996), Research and Development of an International Network for Promoting Grey Literature : A Case Study involving the Use of the Internet. In: *GL'95 Conference Proceedings : Second International Conference on Grey Literature : Grey Exploitations in the 21st Century*, Washington D.C., 2-3 November 1995. – Grey Literature Network Service. – Amsterdam : GreyNet/TransAtlantic, 1996, pp. 247-260. – (GL-conference series, Conference proceedings, ISSN 1386-2316 ; No. 2). – ISBN 90-74854-08-7
5. Farace, D.J. and J. Frantzen ed. (1998), *GL'97 Conference Proceedings : Third International Conference on Grey Literature : Perspectives on the design and transfer of scientific and technical information*, Luxembourg DGXIII, November 13-14, 1997. – Grey Literature Network Service. – Amsterdam : GreyNet/TransAtlantic, March 1998. X, 294 p. ; 30 cm. – (GL-conference series, Conference proceedings, ISSN 1386-2316 ; No. 3). – ISBN 90-74854-17-6
6. Farace, D.J. ed. (1998), *Data Profiles on GreyNet*. – In: *NewsBriefNews*, vol. 7, no. 3, p. 6. – Grey Literature Network Service, GreyNet/TransAtlantic : Amsterdam, 1998. – ISSN 0929-0923
7. Artus, H.M. (2004), Old WWine in New Bottles? Developments in electronic information and communication: Structural change and functional inertia. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp. 1-9. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
8. Blaaij, C. de (2004), Grey Literature from invisibility to visibility: The quest for grey content in the domain of the invisible web. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp. 112-116. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
9. De Castro, P and S. Salinetti (2004), Quality of grey literature in the open access era: privilege and responsibility. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp. 32-39. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
10. Farace, D.J. and J. Frantzen ed. (2004), *GL5 Conference Proceedings : Fifth International Conference on Grey Literature : Grey Matters in the World of Networked Information*, 4-5 December 2003. – Grey Literature Network Service. – Amsterdam : GreyNet/TextRelease, Jan. 2004. – 162 p. – (GL-conference series, ISSN 1385-2316 ; No. 5). – ISBN 90-77484-01-9
11. Gelfand, J. (2004), The Ubiquity of Grey Literature in a Connected Content Context. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. Amsterdam : GreyNet/TextRelease, 2004, pp. 21-26. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
12. Pitoni, I. and D. Macri (2004), GL Systems and Services in the Specific Fields of Vocational Training and Labor Policies in Italy : the ISFOL Case. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp. 124-131. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
13. Siegel, G. (2004), Capturing Academic Grey Literature: Starting at home. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp. 13-20. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
14. Stock, C. and J. Schöpfel (2004), Grey literature in an open context: From certainty to new challenges. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp. 94-105. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
15. Vesely, M., Le Meur, J-Y. and T. Baron (2004), CERN Document Server: Document Management System for Grey Literature in Networked Environment. – In: *GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information"* 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp. 106-111. – (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.

"Knock, Knock:" Are Institutional Repositories a Home for Grey Literature?

Julia Gelfand
University of California, Irvine

Abstract:

Academic and special libraries are eagerly as well as reluctantly joining the bandwagon to participate in institutional repositories. The young and growing collection of the University of California Institutional Repository hosted by the California Digital Library (CDL) contains nearly 5300 documents (see <http://repositories.cdlib.org/escholarship/>). This paper will analyze the contents of that collection in terms of levels of greyness. Content comes from 9 different campuses composing the University of California's Research Units, Centers, and Departments and includes working papers, research content, journals and peer reviewed series. This author has developed a five-point scale that identifies and describes the range of content to conclude the extent that this example of an institutional repository is grey. Institutional Repositories have different collection and review policies and this will be noted. Capturing this content institutionally adds prestige and visibility to resources that without this institutional affiliation may not have peer review, be available digitally and thus remotely, and have perpetual access. A conclusion will be made whether this model of institutional repository supports a new publishing method for renewed life in grey literature.

This conference will focus on a number of different themes that dissect the intersections grey literature has with other forms and formats of information. I will begin the discussion about how institutional repositories have taken on a new role in higher education and scholarly publishing and what implications this has had and may have for grey literature by using the experience of the University of California eScholarship program. An international conference was just sponsored in November 2004 by the Association of Research Libraries (ARL) and the Scholarly Publishing Academic Resources Coalition (SPARC & SPARC Europe) on the next wave of institutional repositories (IR).¹ Meetings of library leaders around the world have taken place over the past few years to launch the emergence of institutional repositories and it is the opinion of this author that it is a way of giving legitimacy to grey literature, even if only scant reference to that has been articulated in the literature or in public forums until now. So knock, knock, someone is really home now.

Clifford Lynch writes about this in early 2003 by stating,

"The development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content and also scaling up beyond ad-hoc alliances, partnerships and support arrangements with a few select faculty pioneers exploring more transformative new uses of the digital medium."²

This powerful and eloquent statement does all to confirm the definition of grey literature by inference and establish a new home for it. Today, still in its infancy, the institutional repository by all accounts is more than a nursery; it is a palace with vast Real-estate as it expands its horizons and hospitality to even more forms of information products and reaffirms its essence and stature in the academic community. Grey literature, as tweaked and redefined at the GL 1997 Conference in Luxembourg is defined as "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."³ For clarity and contrast, the Institutional Repository is best defined by Lynch as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institutions and its community members."⁴

Grey literature differs from commercial publications in that it is not based solely or even principally on an economic model, but rather on a communication model which we also now describe as scholarly communications. This confirms the issues I and many others have written about concerning why grey literature is difficult in collection development and collection management realms. Until electronic publishing and the web was stable it was difficult to identify, obtain and bibliographically describe.⁵ Still, we could benefit from more standards to support even better access.

Two years ago, Roy Tennant of the California Digital Library (CDL) wrote a very concise and informative article on the nuts and bolts of institutional repositories, introducing and explaining the software that ranges from open source to commercial varieties that defines the IR operation. In that short piece, he covers implementation models, keys for access, economic models, subject terminologies and some posting and removal policies. He also has a short concluding paragraph entitled, "From grey to black and white," where he concludes, "They provide much better access to a literature than has ever previously been possible and should be a no-brainer for most academic institutions."⁶

This reference to grey literature and the remainder of the context of Tennant's paper leads me to conclude that new learning communities have formed or will form as the result of institutional repositories being established and maturing with the depth and scope of their content. It is my speculation that the IRs will really make a difference in higher education because the content has increased legitimacy, offers simpler and better access, has gone through peer review, promotes local institutional research and output and is in the format our contemporary users prefer.

The literature on learning communities goes back nearly two decades now and incorporates lots of principles from the early work of Ernest Boyer, Peter Senge and many others. We must ask ourselves what types of learning communities are most relevant to institutional repositories and I believe a consensus would emerge that cross-curricular, purposeful, electronic, primary and secondary would all apply.⁷ The Lenning and Ebbers volume, *The Powerful Potential of Learning Communities*, also explores the benefits of learning communities to building learning environments where the faculty benefits include: "diminished isolation, a shared purpose and cooperation...among colleagues, increased curricular integration, a fresh approach to one's discipline and increased satisfaction with their student's learning."⁸ These attributes are parallel or perhaps even synonymous to criteria that librarians apply to the selection principles they use to guide their decisions about what library collections will contain, what materials will be retained where, what format of a product will be chosen and other decisions that impact the future of learning communities and libraries on both a macro and micro level, and information delivery and access as it applies to a wide spectrum of understanding and practice.

If nearly all learning communities have "two things in common with one being shared knowledge and the other, shared knowing"⁹ then Gardner's list of what a learning community does fits almost seamlessly into our exploration of institutional repositories as a rightful and purposeful place for grey literature. He suggests that learning communities:

1. Incorporate and value diversity
2. Share a culture
3. Foster internal communication
4. Promote caring, trust and teamwork
5. Involve maintenance processes and governance structures that encourage participation and sharing of leadership tasks
6. Foster the development of young people and
7. Have links with the outside world.¹⁰

Returning to the main thrust of my paper of how grey the California Digital Library's eScholarship collection is requires a few steps. My observation and participation in the program is limited to three vantage points: as an affiliated librarian at one of the 10 University of California campuses where I direct users to resources and direct faculty to publishing options, and finally as an independent user. In the first and third roles, I am searching for information that may not be in standard or traditional resources, meaning published books, journal collections, conference proceedings or such packages, where I use finding aids such as indexing and abstracting services or databases. The second role is a rather new departure for librarians. In recent years we have become increasingly comfortable with different facets of scholarly communication and are familiar with alternative options within the accepted publishing spectrum and now with new products and electronic mediums. We are also committed to promoting how economically and spatially unsustainable earlier models of scholarly publishing have become. In addition we have a commitment to digital preservation, open access, authors' rights, cost containment and many other creative and intellectually motivated goals for scholarly publishing. UC Librarians encourage and promote the eScholarship activities and repository as a place to capture and retain the research and scholarship that is conducted by our leading faculty.

The new modes of scholarly publication within the University of California (UC) eScholarship program include: **institutional repositories** that promote pre-publication of materials and contain peer-reviewed content; **web-based publications** of digitally reformatted content and electronic editions of academic monographs of interest to both scholarly and general-interest readers. In addition, there are numerous partnerships with librarians, scholars, publishers, a wider information industry community where a special eScholarship repository has been created to support a fuller range of scholarly output from pre-publication materials to journals and peer-reviewed series, by offering the University of California departments and units direct control of publishing.¹¹

You may ask and ponder why create eScholarship and an institutional repository for one university. Please remember that the University of California has 10 campuses participating and the Repository reflects a full spectrum of publishing activity from reports, peer-reviewed content, edited volumes,

journals, pre-prints and in January 2005 will include post-prints. Membership is by the academic research unit and/or department which serves as "gatekeepers" of the content and where editorial and administrative functions are distributed.¹² However interrelated eScholarship and institutional repositories are, I am just going to concentrate on the IR component for our discussion today.

The repository is searchable by:

- Campus
- Research unit, center or department
- Journals and peer-reviewed series
- Seminar series

Usage counts as of December 1, 2004 indicate that the Repository had 837,339 downloads to date and the most recent week experienced 20,235 downloads. John Ober in his presentation at the ARL/SPARC Institutional Repositories Conference documents patterns of downloads and participation from academic units.¹³

As the repository grows to currently reflect more than 5300 items I became curious about what the common and divergent elements are and this is what led me to dissect the content to determine the degree of grey literature that it contains, especially since the majority of use of the IR is from outside the UC system. Placing the cart before the horse, let me share with you that this was a reasonably easy analysis because all the pieces fell into place without too much scattering. The method was a simple, quasi social science effort of describing content based on a palette of colors that represented certain matches. Content could be coded multiple times if appropriate, although only 37% of my sample had that possibility. It is important to understand the flexibility and context of the Repository Policies before any conclusions can be made because an interesting comparison for future studies is whether they are indeed sufficiently flexible. Currently, they include:

- Who can join
- Whose papers can be included in the Repository
- Appropriate submissions
- Peer-reviewed series
- Seminar series
- Removing a paper
- Author review & agreements
- Copyright¹⁴

Since authors retain the copyright for all content posted in the repository and the eScholarship initiative features a non-exclusive right so that the author is free to use the content in multiple way. This may be new for some faculty and guidelines such as the following passage may help inform specific practices:

If a working paper is published in a journal—either in the same form or, more commonly, in revised form—many journals allow the working paper to continue to be made available, especially when it is for educational/scholarly noncommercial use. Unfortunately, some journals do require that the working paper be removed. Others grant exceptions for something like the eScholarship Repository; they just need to be asked. It is up to the faculty member to check the terms of their agreement with the journal to see what is allowed. Individual journal policies vary widely. The RoMEO Project (Rights METadata for Open archiving) has compiled a list of many journals' "Copyright Policies" about "self-archiving."

The Repository Benefits are equally important and they may be even more relevant to how we consider grey literature. Selective benefits include:

- Free to contribute for all University of California affiliates
- Promising alternative
- Increased visibility
- Usage reports
- eMail notification alerts to readers & users
- Permanence
- Global accessibility via the Open Archives Initiative (OAI)
- Ability to upload associated content
- Institutional identity
- Sophisticated searching
- High quality participants¹⁵

Since the content is available for all search engines to crawl, discover, and make available to their users, access is extended over the broadest range so the indexing and abstracting of content is among the least restrictive. This expands access beyond any previous distribution model that any form of grey literature has experienced in its history. Each search engine provides the output using different algorithms reflecting high use, relevancy, currency, etc. Obviously, part of this is due to the potential of electronic publishing formats and new assignments of authors' rights, but the blending of content into different types of information products creates some of the richest searching sources for relevant information. The recent release of Google Scholar (<http://www.scholar.google.com>)

enhances access and offers different forms of competition to users and will probably include the eScholarship Repository in its output soon.

The palette of colors reflects the following spectrum and was coded to items in the repository according to the following assignments with greyness sliding from each color and blending in the middle. This rather simple methodology assigned a numerical value for each color or factor on a range of 1-5 with one being low and 5 high. This demonstrated how many examples in the repository that were reviewed had strong indicators for linking, environmental, publishing, collaboration and interdisciplinary factors.

Black - > **Linking Factors** - demonstrates relationships to author's other works (backward -> forward), to institutional colleagues, to other content and secondary sources that track citation histories, etc

Red - > **Environmental Factors** - reflects the complexity of initial acquisition, description, funding, troubleshooting and support for persistent and perpetual access tracking versions and usage, etc

Blue - > **Publishing Factors** - even though the effort is to reduce time to publication with the attributes of peer review, capturing older work that has never been shared or previously released is another goal as is the overall publishing process. In the future, citation analysis may be a part of this category.

Yellow - > **Collaborative Factors** - can be used in multiple functions, for instance in classroom teaching, scholarship, repackaging, etc

White - > **Interdisciplinary Factors** - promotes and incorporates the spirit of new and emerging work from multiple subject areas, and may include conference and seminar content, etc.

In June 2004 I reviewed a randomized sample of content in the following content areas: 50 examples of the research units/departments and 50 examples from the Seminar Series. Due to the 37% of multiple codings the total sample included 137 submissions which is only 2% based on today's inventory or 8% in June. Sample size is very modest but the highest matches were for collaborative factors, followed by interdisciplinary factors, environmental factors, publishing factors and then linking factors.

Linking Factors - 7% or 9.59 items

Environmental Factors - 18% or 24.66 items

Publishing Factors - 21% or 28.77 items

Collaborative Factors - 25% or 34.25 items

Interdisciplinary Factors 29% or 34.25 items

Total = 100 % or 137 items

The range reflected more about grey literature than it did about other descriptive elements. My hypothesis that as the Repository grows at its rapid rate and there are more citations attributed to its content, and that the colors will blend even more and become less grey but more white, even with linking a problematic factor for a longer while. The reason for this is that linking depends on busy academics to provide the information and add more of their earlier work to offer a spectrum of work and to encourage other colleagues to contribute to the IR. Maturity in this category is more dependent on external factors than any of the others.

To conclude, we can say that the factors defined as linking, environmental, publishing, collaborative and interdisciplinary each describe grey literature contained in institutional repositories. There are longstanding issues including:

- transience of grey literature
- maturation of the repository
- timely publishing
- access
- standards
- multiple formats

and other areas that need more attention. The players and partners will change and the content will not only multiply but absorb new technologies and discoveries to meet user expectations. The challenges that will remain for the foreseeable future include multiple formats, how to scale content, keep it easy, changing technologies, identification and access, need for a mixture of expertise to implement projects and leveraging of collective investments, communication and promotion of the content¹⁶ But we know that Grey Literature will not be homeless again as long as institutions continue to exploit the possibilities and merits of building and refining the institutional repository concept. No longer is Grey Literature at risk, it will be in a good collegial neighborhood, and be sought after instead of being a weak commodity in a chain of information products that was previously inhospitable to it. Libraries are happier because the organizational structure of the institutional repository remains academic and scholarly and there is no fear of being lost. The palette of factors will become brighter in each category and fading grey will be a part of the past. The proliferation of institutional repositories and high usage and downloading is already testimony to this transformation. Knock, Knock, someone is certainly home.

Author Information

Ms. Julia Gelfand, Applied Sciences & Engineering Librarian
Science Library 228, Research & Instructional Services
University of California, Irvine Libraries (UCI)
P.O. Box 19556
Irvine, CA 92623-9556
USA

EML: jgelfand@uci.edu
URL: <http://www.lib.uci.edu>

Cited References

- ¹ See <http://www.arl.org/sparc/meetings/ir04/ir04speak.html>
- ² Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," ARL Bimonthly Report 226, February 2003. <http://www.arl.org/newsltr/226/ir.html> also published in *Portal: Libraries and the Academy*, 3, 2, 2003:327-336.
- ³ Dominic J. Farace, "STS Publisher/Vendor Relations Conference Discussion Group Conference Report," published in ISTL, Winter 1998. <http://www.istl.org/98-winter/conference4.html>
- ⁴ Lynch, 327
- ⁵ Julia M. Gelfand, "Academic Libraries and Collection Development: Implications for Grey Literature," Publishing Research Quarterly 13,2, Summer 1997: 15-24 and "Grey Literature Poses New Challenges for Research Libraries," Collection Management 24,1/2, 2000:137-148, and Gelfand and John L. King, "Grey Market Science: Research Libraries, Grey Literature and Legitimization of Scientific Discourse in the Internet Age," in *Socioeconomic Dimensions of Electronic Publishing Workshop Proceedings*, 1998:115-120. Paper presented at IEEE Conference, Santa Barbara, CA, April 25, 1998.
- ⁶ Roy Tennant, "Institutional Repositories," *Library Journal*, September 15, 2002. <http://www.libraryjournal.com/iondex.asp?layout=articlePrint&articleID=CA242297>
- ⁷ Oscar T. Lenning and Larry H. Ebbers, *The Powerful Potential of Learning Communities: Improving Education for the Future*. ASHE-ERIC Higher Education Report 26 #6. Washington, DC: The George Washington University Graduate School of Education and Human Development, 1999:iii.
- ⁸ *Ibid*, iv.
- ⁹ Vincent Tinto, "Colleges as Communities: Taking Research on Student Persistence Seriously," *Review of Higher Education* 21, 2, 1998:171 as referenced in Lenning and Ebbers, 10.
- ¹⁰ John W. Gardner, "Building Community," *Kettering Review*, 1989: 73 as referenced in Lenning and Ebbers, 10.
- ¹¹ California Digital Library eScholarship website at <http://www.cdlib.org/prgorams/escholarship.html>
- ¹² Catherine Candee, "eScholarship," Presentation made at the Society for Scholarly Publishing Annual Meeting, June 4, 2004, San Francisco.
- ¹³ John Ober, "IR Content & Service Expansion: the Case of eScholarship," Presentation at Institutional Repositories: the Next Stage Conference Nov 18-19, 2004, Washington, DC. <http://www.arl.org/sparc/meetings/ir04/presentations/ober.html>
- ¹⁴ See <http://repositories.cdlib.org/escholarship/policies.html>
- ¹⁵ For complete list see <http://repositories.cdlib.org/scholarship/benefits.pdf>
- ¹⁶ Adapted from Lucia Snowhill, "Working with Government Information System-wide," Presentation made to UC Conference on Government Information, October 28, 2004.

In addition to the cited references these resources also contributed to this paper:

Carol A. Hughes, "eScholarship at the University of California: A Case Study in Sustainable Innovation for Open Access," *New Library World*, 105, 1198/1199, 2004: 118-124.

Institutional Repositories: Revealing Our Strengths. ARL/OLMS Webcast cosponsored by SPARC and CARL, June 10, 2004.

Marie E. McVeigh, "Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns." White Paper from Thomson Corporation, 2004. Available at <http://www.thomsonisi.com/media/presentrep/essayspdf/openaccesscitations2.pdf>

The opinions expressed in this paper are solely those of the author and do not imply agreement or endorsement of the CDL staff.

Sharing Grey literature by using OA-x¹

Elly Dijk

Netherlands Institute for Scientific Information Services, NIWI-KNAW

Abstract

As part of the Dutch DARE (Digital Academic REpositories) programme, NIWI-KNAW is participating in various projects to enlarge open access to Dutch scientific output (including Grey Literature). The Open Source web technology that will be used for harvesting is based on i-Tor, Tools and technology for Open Repositories, developed by NIWI (the Netherlands Institute for Scientific Information Services).

The subject of this lecture is the initiative that NIWI-KNAW has taken for developing the OA-x protocol, a modular extension of the OAI protocol, OAI-PMH (Protocol for Metadata Harvesting). We shall discuss the advantages of OA-x and for what projects it will be used.

In the protocols of the Open Archives Initiative (OAI) currently in use, information is shared by providing metadata of digital files (data providing) that can be read in by someone else (data harvesting). A URL is used to refer to an object in an external site (often a repository).

In certain cases, one needs to go further than just sharing metadata. Certainly in the case of so called collaboratories, it should be possible to transfer the objects themselves from an external website (or repository) to one's own site. And conversely, it should be possible to upload objects to an external site. Even if only the browsing of objects is required, it is necessary to get to the original document in order to be able to index it.

The OA-x project has been set up to enable researchers and administrators of (digital) archives to be able to unlock, edit, supplement, combine and archive metadata and data (objects) in digital repositories. A protocol for harvesting and uploading objects has been developed in this project. There are also several implementations available: OA-x within a CMS, OA-x as extendable OAI data and server provider, and OA-x as repository filler. We have opted for similar names of verbs as are used in OAI-PMH.

The advantages for authors and administrators of (digital) archives are great. It is possible to place articles or other publications on one's own website as full text and easily export them to a repository such as an institute repository. It is also possible to use OA-x to upload publications to electronic journals (e.g. Studies in Mycology) or to a central address where a grey publication will be produced.

With the aid of the i-Tor technology, it was already possible to index PDFs on one's own website as full text and make them searchable via Google. Thanks to OA-x, it is now possible to index PDFs (or other text files) on external sites as full text too.

In collaboratories, it is not uncommon to use collections of images that are split over various sites. With an OA-x implementation, it is possible to make a collection of thumbnails of the images in these distributed collections in one place (as if the images were collected on one site only).

An example of a collaboratory is E-laborate, a virtual joint venture in the alpha and gamma disciplines. OA-x is used by E-laborate to upload datasets to subject-based repositories.

OA-x can also be used to make a (national) electronic depot. Objects can be sent to such a depot or archive from institutional repositories. OA-x makes it possible to not only send the object but also multiple datasets. It is even more important that information about the technical data (e.g. in what version of PDF the object was created) can be sent along with it. These data are of essential importance to a depot because they can be used to see whether an object (and the format of the object) has remained unchanged.

Introduction

As part of the Dutch DARE (Digital Academic Repositories)² programme, the Netherlands Institute for Scientific Information Services (NIWI)³ – an institute of the Royal Netherlands Academy for Arts and Sciences (KNAW)⁴ – is participating in various projects to enhance open access to Dutch scientific output, including "grey" literature. DARE is a joint initiative by the Dutch universities to make their research data accessible in digital form. As well as KNAW-NIWI, its other participants include the Netherlands

Organisation for Scientific Research (NWO) and, from the conservation perspective, the National Library. The programme began in 2003 and will last until 2006. DARE is being co-ordinated by SURF, an ICT partnership set up by the Dutch universities.

The original aim of DARE in 2003 was to establish institutional repositories of academic output at all Dutch universities and KNAW and NWO institutes. These would contain working papers and pre-prints, dissertations and theses, research reports, datasets, contributions to congresses and multimedia presentations. These repositories form a distributed network: they can be searched either collectively or individually.

The web technology developed by NIWI for use as a content management system and service provider for DAREnet⁵, the DARE website, is called i-Tor⁶ (Tools and Technology for Open Repositories). i-Tor has also contributed towards a number of DARE projects at individual universities, as well as to developing a service provider to harvest digital material from the universities and a data provider for the KNAW repositories and others. These results were presented at a meeting in January 2004, after all the universities and a number of KNAW and NWO institutes has set up at least one repository which could be searched through DAREnet. These repositories contain a huge range of different material: publications, including grey literature; films; audio fragments; and so on. In total, the repositories now contain more than 20,000 digital files.

The aim of the DARE community in 2004 is to add to these repositories, primarily by increasing scholars' involvement in the programme.

Also in 2004, NIWI began the OA-x project. Financed by DARE, this has been established to enable researchers and the managers of digital and other archives to retrieve, edit, add, combine and archive both metadata and objects from digital repositories. As part of the project, a new protocol for the harvesting and uploading of objects has been developed. OA-x, which is also available as a plug-in for i-Tor, is open source technology and can be regarded as a modular extension of the Open Archives Initiative Protocol for Metadata Harvesting (OAI/PMH).⁷

This paper looks first at the i-Tor open source web technology which formed the basis for the development of OA-x. The four areas covered by i-Tor – content management, archives, collaboratories (digital workplaces) and e-publishing – are all addressed. We then provide details of the OA-x project, describing its background, the OA-x protocol and possible applications. The paper then addresses DIDL and METS, before finally mentioning those projects currently using OA-x.

1. i-Tor: a new open source web technology⁸

The i-Tor web technology is being developed by NIWI, in collaboration with various universities and KNAW institutes in the Netherlands as well similar bodies in Germany and Belgium. i-Tor can be regarded as a toolbox for use when creating a website, collaboratory, information portal, repository, database retrieval system, and so on.

1.1 Content management system (CMS)

i-Tor was developed originally as a content management system (CMS) to upgrade the NIWI website. The old site consisted of static HTML pages maintained by webmasters. Thanks to i-Tor, it is possible for non-specialist staff at the institute and elsewhere to keep their own information up to date. Full-text searches are possible in all data, be it web texts, PDF documents or database records. And the content of databases is accessible by search engines like Google. i-Tor is used in the construction of internet, intranet and extranet websites. Access – reading and writing rights – is governed by permissions. In terms of such aspects as layout, i-Tor can be adapted to suit the user organisation. And it is flexible – for example, there are no coercive limitations in workflow.

This CMS forms the basis for further development of the other areas covered by i-Tor. In partnership with various NIWI departments, the development team is enhancing and simplifying the CMS to ease database retrieval and improve the search functionality. These updates will be completed within a few months.

As well as the NIWI, about 30 other institutes and projects – both national and international – have now (in October 2004) constructed their own websites with the aid of i-Tor technology.⁹

1.2 Archives

NIWI is also developing i-Tor in other areas. Its Department of History, for example, is a participant in the European Visual Archives (EVA) project.¹⁰ This is working on the searchability of archives containing digital images, which generated a request to develop i-Tor into an OAI service provider for retrieving available metadata and an OAI data provider, making data available in open archive form. The OA-x project described below also falls into this i-Tor category.

As well as the EVA project, the OAI service provider and data provider are used by the DAREnet mentioned earlier. And NIWI is also working with the German Centre for Polar and Marine Research¹¹ at the Alfred Wegener Institute to create a kind of large-scale service provider.

1.3 Collaboratory

The word "collaboratory" is a combination of "collaboration" and "laboratory". i-Tor allows the creation of a virtual research space for academics working at different physical locations. That space is the collaboratory. It can hold large amounts of information – texts, audio, raw data, databases, video and so on – in various formats (including PDF and XML) for sharing in a straightforward and consistent way.

Functions like authorisations for reading and writing rights, weblogging, RSS feeds, and discussion lists and version management are essential to a collaboratory. All are available in i-Tor, or are being developed for it.

One example of such a collaboratory is E-laborate¹², a partnership in the humanities and social sciences. The aims of this project centre on opening up opportunities to share and collaborate on textual material and datasets. The latter is the subject of a subproject, X-Past. As a test case for the Textual Material subproject, the E-laborate Steering Group has chosen the Dutch historical and cultural journal *Vaderlandsche Letteroefeningen*. This is being used to develop electronic tools for co-operation related to texts and textual material. Those tools will also be applicable to other textual material.

1.4 E-publishing

Other i-Tor functionalities are being developed in the area of e-publishing. At the request of another KNAW institute, the Fungal Biodiversity Centre (CBS), NIWI has created a publication tool for the production of a digital journal – in this case, *Studies in Mycology*.¹³

As part of the DARE programme, a "similarities" function has been developed for three Dutch universities. This is actually a plagiarism scanner, which can search students' dissertations for stolen work.¹⁴

Another function, still under development, will enable peer review.

2. The OA-x project

2.1 Background

The OA-x project was established to enable researchers and the managers of digital and other archives – including those containing grey literature – to retrieve, edit, add, combine and archive both metadata and data (objects) from digital repositories.

Under the standard Open Archives Initiative protocols currently in use (OAI-PMH), information is shared by supplying metadata from digital files (data providing) in a form which can be read by others (data harvesting). A reference may then be provided, through a URL, to an external site – often a repository.

In certain cases, however, users wish to go further than simply sharing metadata. Certainly within the context of a collaboratory, it is preferable that the objects themselves be transferable from an external website or repository to the user's own site. And, conversely, users want to be able to upload objects to an external site.

NIWI-KNAW has conducted the OA-x project with the help of a subsidy from DARE. The project was designed to produce a protocol for the harvesting and uploading of objects. At the same time it also addressed the digital durability of the objects stored. For objects to be retained and accessible in the long term, it is not enough to be able to harvest the metadata alone. The objects themselves also have to be harvestable. Yet the current OAI-PMH is inadequate for that. OA-x, on the other hand, makes it possible

to transfer metadata together with the associated objects. And it allows the user to check that the metadata or the object or the combination of the two has been received correctly.

2.2 The protocol

In developing OA-x, it was decided that it should be easy to implement for existing OAI-compliant repositories and that everything should be done in as transparent a way as possible, building upon what has already been achieved with OAI-PMH and i-Tor. So OA-x has been designed as a modular extension of OAI. Developed as a plug-in for i-Tor, amongst other forms, OA-x, is open source technology. Some of its features have been implemented within i-Tor, thus creating a proof of concept with which the specific properties of OA-x can be demonstrated.

OA-x has been developed as an extension of the OAI/PMH protocol.¹⁵ It has also been decided to give its commands names similar to those in OAI/PMH.

The following are the most important commands.

- **GetObject(s).** Used to harvest objects. The plural **GetObjects** command has been created for bulk processing.
- **PutObject(s).** Used to upload objects to an external repository. The plural **PutObjects** command has been created for bulk processing.
- **Checksum.** A unique number, generated using a special algorithm, which can be assigned to an object.

The CheckSum command works as follows.

- Data provider calculates checksum and data.
- Service provider harvests data.
- Service provider requests the relevant checksum from the data provider.
- Service provider also calculates the checksum of the harvested data.
- If the checksums match, everything is in order. In not, there is something wrong with the harvested data.

NB. OAI itself has no checksum mechanism.

The similarities and differences between OAI-PMH and OA-x shown in the table below.

OAI-PMH	OA-x
Identify	Identify
ListRecords	ListRecords
ListSets	ListSets
ListIdentifiers	ListIdentifiers
ListMetadataFormats	ListMetadataFormats
GetRecord	GetRecord
	GetObject
	GetObjects
	PutObject
	PutObjects

Logically, the OA-x object closely resembles an OAI record. But they differ in the following respects.

- The header with identifier and datastamp.
- The metadata (DC or any other pattern).
- The body (a PDF, image or any other form of binary).

Clearly, this latter point is where OA-x represents an enhancement. An example of an OA-x object is given in Appendix 1.

2.3 OA-x applications

OA-x considerably broadens the possibilities available to users, thus significantly boosting the potential for virtual collaboration. The benefits for authors and the managers of digital and other archives are enormous.

Some important applications of OA-x are listed below.

- **Use of the GetObject command**

- Within i-Tor, it was already possible for users to provide the PDFs on their own site with full-text indexing to make them searchable by Google. Thanks to OA-x, the same can now be done for PDFs or other text files on external sites.
- Collaboratories often make use of collections of images which are spread across several sites. An OA-x implementation makes it possible to produce a collection of thumbnails from these distributed collections for placement on one site. This makes it appear that all the images are in one place.
- OA-x can be used to upload datasets to subject-based repositories.
- OA-x can also be used to integrate full-text objects.
- Analytical tools – for grammatical or textual analysis, for example – can be used even when documents are distributed across several sites.
- At large academic institutes, news bulletins are often generated in different places. The OA-x protocol can be used to bring together these bulletins, or parts of them, at a central point.
- Establishing relationships between objects – for example, by using tools such as the "similarities" function mentioned above – is easier with OA-x than in previous protocols.

- **Use of the PutObject command**

- It is easy to upload articles and other academic information sources to the repository of another institution.
- An export from an existing application or database to an external repository can be generated. At present, this is only possible between i-Tor sites.
- Users can place full-text articles or other publications on their own website and at the same time easily create an export to a repository – for example, that of their institute.
- OA-x can also be used to upload publications to electronic journals or to a central address where a grey publication is produced.

- **Use of CheckSum**

- A checksum is a unique number, generated using a special algorithm, which can be assigned to an object. To check that an archive holds the original object – which is important when working in a collaboratory environment, for example – such a number is vital. Even the tiniest adjustment to the data will change the checksum irrevocably and thus alert the researcher.
- Naturally, such a checksum can be extremely useful in such areas as version management and the verification of electronic publications.

OA-x and the electronic depot: a potential special application

OA-x could contribute to the creation of a national electronic depot or archive, using the PutObject command to submit objects drawn from institutional repositories. Not only can the object be sent, but also multiple metadata sets. Even more importantly, technical information – such as which version of PDF the object was generated in – is sent as well. These details can be vitally important to the depot, since they are used to check whether the object or its format have been changed. This is done using CheckSum. Transferring the object to a different environment, such as a new version of PDF, will automatically generate a different CheckSum.

If an institution wants to retrieve objects – its own or other people's – from the electronic depot, it can use the GetObject command.

2.4 DIDL, METS and OA-x¹⁶

There is a long-standing wish to be able to retrieve actual objects, not just their metadata. And two standards have already been developed to do this: DIDL and METS.

MPEG 21 DIDL (Digital Item Declaration Language)¹⁷ is a metadata format for any digital object and contains all the elements needed to harvest various objects automatically. But it does require agreements to be reached about a special metadata format for the dissemination of a resource. One solution to this limitation would be to make that dissemination part of the protocol. The OAI Working Group could rule on that. In consultation with service and data providers, it would then decide what such an extension should look like. But it is going to be some time before a consensus is reached on this issue.

Thus far it has proven practically and, in particular, organisationally difficult to make repositories MPEG 21 DIDL-compliant. For it to be usable in developing an electronic depot, the protocol would at the very least have to be enhanced to include the PutRecord command.

Another possibility is to use METS (Metadata Encoding and Transmission Standard)¹⁸. This is a metadata model in which it is possible to include comprehensive information about a particular resource. In terms of function, METS is comparable with the Dublin Core: it is intended as a standard for data exchange. But METS does not state which protocol is used to share that data. Any protocol can be used: OAI, SOAP and so on. METS' users include the US Library of Congress and its scope in terms of resources is reasonably broad, encompassing text, audio, video, photography and so on, but not datasets.

Since you can use it as XML, METS combines excellently with OAI. It is relatively easy to send an METS XML record using OAI. However, METS is no OA-x. It is not itself responsible for the sharing of resources, any more than DIDL is. Both METS and DIDL are metadata standards in which you can *either* provide links to the resources *or* incorporate the resources themselves in their entirety. The mechanism which actually ensures that the resource reaches the user is not specific in either standard. The two possible scenarios can be summed up as followed.

1. The METS/DIDL description contains a link to the resource, and this must be followed in order to download the actual resource.
2. The METS/DIDL description contains the actual resource, which first has to be "cut" from that description and saved separately before it can be used.

In this respect, OA-x goes further. Whereas METS and DIDL contain either the resource or a link to it, so that you always receive the resource through OAI if you opt for them, OA-x actually draws a distinction between harvesting the metadata and harvesting the resource. To do this it has added separate commands (GetObject) to the OAI protocol, which METS and DIDL have not.

2.5 Current use of OA-x

Within the DARE community, the development of OA-x has prompted a debate as to whether it is actually going to be used in DARE. Contact has been established with the OAI community and the various options – OA-x, DIDL and METS – are being considered.

As mentioned earlier, however (see 2.3), OA-x has already found a number of applications. The E-laborate collaboratory, a partnership in the humanities and social sciences, is using it to update datasets into subject-based repositories as part of the X-Past subproject. The Bibliography of Dutch Language and Literature (BTNL) is also using OA-x, to integrate full-text objects. A third application is the upload of publications to electronic journals, as is happening at *Studies in Mycology*.

OA-x has been designed as a modular extension of OAI. Some of its features have been implemented within i-Tor (Tools and Technology for Open Repositories), thus creating a proof of concept with which the specific properties of OA-x can be demonstrated.

Full details of OA-x can be found on the i-Tor website, http://www.i-tor.org/oa_x, together with comprehensive descriptions of DIDL and METS.

Literature

- Arjan Hogenaar. *Eindrapportage Project OA-x* ("Project OA-x End Report"). NIWI-KNAW, 21/04/2004.
- Marc Evers, *i-Tor naar een modulaire plugin-architectuur met Eclipse* ("i-Tor: towards a modular plug-in architecture with Eclipse"). In the electronic newsletter *Reposi-Tor*, no. 5, September 2004.
- Rutger Kramer; *MPEG-21 DIDL Approach; OA-x Approach* (http://www.i-tor.org/oa_x/retrieving_objects). July 2004.

Appendix 1: Example of an OA-x object (PDF object itself has been shortened to make the example not too long)

```
<?xml version="1.0" encoding="UTF-8" ?>
<OA-x xmlns="http://oax1.cq2.org/OAX/0.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://oax1.cq2.org/OAX/0.1/ http://oax1.cq2.org/OAX/0.1/OA-x.xsd">
<responseDate>2004-02-06T15:47:02Z</responseDate>
<request verb="GetObject" identifier="bob">http://oax1.cq2.org/nl/story4test</request>
- <GetObject>
  - <object>
    - <header>
      <identifier>bob</identifier>
      <timestamp>2004-02-06T15:47:02Z</timestamp>
    </header>
    - <metadata>
      - <oai_dc:dc
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:creator>thijs</dc:creator>
          <dc:type>pdf</dc:type>
          <dc:title>bob</dc:title>
          <dc:date>2004-02-06T15:31:50Z</dc:date>
          <dc:contributor>thijs</dc:contributor>
          <dc:format>application/pdf</dc:format>
        </oai_dc:dc>
      </metadata>
    - <data>
      - <itor:itor
        xmlns:itor="http://oax1.cq2.org/"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://oax1.cq2.org/OAX/0.1/itordata.xsd">
          <itor:originalfilename>bob.pdf</itor:originalfilename>
          <itor:contentbytes encoding="base64"
            mimetype="application/pdf">JVBERi0xLjIgDSXi48/TDQog...
            cmVmDTM1NjU0DSUIRU9GDQ==</itor:contentbytes>
          </itor:itor>
        </data>
      </object>
    </GetObject>
  </OA-x>
```

Author Information

Ms.drs. Elly Dijk,
Policy Officer
Netherlands Institute for Scientific Information Services, NIWI-KNAW
P.O. Box 95110
1090 HC Amsterdam
The Netherlands

Elly.Dijk@niwi.knaw.nl
www@niwi.knaw.nl

Notes

¹ The author would like to thank Rutger Kramer (NIWI-KNAW) and Laurents Sesink (DENK-KNAW) for reviewing this paper critically.

² For more information about the DARE programme, see <http://www.surf.nl/themas/index2.php?oid=18>.

³ For more information about NIWI, see <http://www.niwi.knaw.nl>.

⁴ For more information about KNAW, see <http://www.knaw.nl>.

⁵ DAREnet has been active since January 2004. On this website, <http://www.darenet.nl>, it is possible to search the various institutional repositories.

⁶ For more information about i-Tor, see <http://www.i-tor.org>. The i-Tor source code is available on SourceForge, at <http://sourceforge.net/projects/i-tor>.

⁷ For information about the OAI-PMH protocol, see <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

⁸ The i-Tor web technology is made up of various open source components and is based upon open standards (OAI). Built in Java, it is platform-independent and uses a MySQL database. i-Tor is Linux-based and contains software developed by the Apache Software Foundation. It can be implemented as an independent package. Between now and January 2005 the i-Tor software is being converted to a modular plug-in architecture. This is being done using the Eclipse framework originally developed by IBM. For more information, see <http://www.eclipse.com>

⁹ See the i-Tor website, <http://www.i-tor.org>, for a list of those sites built using i-Tor.

¹⁰ For more information about the EVA project, see <http://www.eva-eu.org>.

¹¹ The Alfred Wegener Institute for Polar and Marine Research is a member of the Helmholtz Association of German Research Centres. For more information, see <http://www.awi-bremerhaven.de>.

¹² For more information about E-laborate, see <http://www.e-laborate.nl>.

¹³ The home page of the electronic publication *Studies in Mycology* is <http://www.studiesinmycology.org>.

¹⁴ The DARE project Dissertations Online, a joint initiative by the University of Twente, Erasmus University Rotterdam and the University of Groningen, has its own i-Tor website, <http://www.scripts-online.nl>.

¹⁵ For more information about this protocol, see <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

¹⁶ For more information about OA-x, DIDL and METS, see http://www.i-tor.org/oa_x/retrieving_objects/.

¹⁷ For more information about MPEG 21, see <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>. There are several internet sites on DIDL.

¹⁸ For more information about METS, see <http://www.loc.gov/standards/mets/>.

Wallops Island Balloon Technology: Can't see the Repository for the Documents

**Andrea Japzon and Nikkia Anderson
Information International Associates, Inc.
NASA Goddard Space Flight Center**

Abstract

Since the Wallop's Balloon Technology documents repository began approximately 9 years ago, the Goddard Library has become increasingly involved in developing digital archiving capabilities. The Library developed the Digital Archiving System (DAS) which is a prototype infrastructure for creating a combined metadata repository that allows metadata for heterogeneous digital objects to be searched with a single search mechanism and presented in a single results page. With this, the opportunity has been presented to expand the usability of the print repository. The Balloon Technology documents relate only to the specific subject of construction of scientific balloons and at the current time number over 4,300. The documents exist primarily in paper format and are organized according to the accession number. The project is currently at a crossroads where decisions will have to be made regarding the reorganization of the database from many different perspectives. An assessment of the project was conducted to determine future direction. An assessment survey was created using the Knowledge Management Assessment Tool (KMAT) from the American Productivity & Quality Center and from the recommendations that The Scholarly Publishing & Academic Resources Coalition (SPARC) put forth in "The Case for Institutional Repositories: A SPARC Position Paper. Survey participants agreed to move forward with project by scanning the documents, mapping existing database records to the current metadata elements, seeking copyright permissions, and forming a joint committee of balloon scientists and librarians. All have agreed on the importance of digitizing this collection to the balloon science community. Further, it was agreed that once complete, the addition of the balloon documents to the DAS (an institutional repository) could serve as a model for other NASA and/or government projects trying to organize, collect and preserve specialized knowledge that manifests largely in grey literature.

History of the Library and the Database

The Balloon Technology Library and Database was created to provide a single repository for scientific ballooning literature and data. Grey and commercially published literatures comprise the collection of technical reports, working papers, proceedings, and journal articles. The primary impetus for creating the library and database was the fact that much of the literature relating to balloon technology is grey literature and therefore not easily found, let alone found in one place. The Balloon Technology documents, currently numbering over 4,300, relate specifically to the construction of scientific balloons. The subject matter covered includes balloon theory, balloon physics, design, operations, performance, facilities, testing, materials, fabrication, quality control, failure analysis and history [1].

The Balloon Technology collection project started in November of 1993 continues to exist today with the sustained involvement of Wallops Flight Facility (WFF) balloon scientists and Goddard Space Flight Center (GSFC) librarians. WFF is a part of GSFC but located three hours south of Greenbelt, Maryland in Wallops Island, Virginia. The WFF mission includes the performance of research in the areas of sub-orbital and small orbital payloads. Balloon science experiments are a significant part of this mission. From the beginning of the project, the intention of the Balloon Program Office (BPO) at WFF was to start a library that would become the World's Archival Center for Scientific Balloon Technology [2]. To that end, project participants set out to and accomplished the major goals of collecting to the greatest number possible of existing balloon technology documents and creating a searchable database. With these accomplishments in hand, now is the time to increase the availability of this work to balloon science researchers.

A great debt is owed to Jim Winker, balloon scientist and de facto librarian, for performing the yeoman's duty in reaching the goals set out by the BPO. In the field of librarianship, perhaps the greatest value placed on information is its authority. In the creation of this collection, Jim Winker's life experiences provide that authority. He possesses extensive historical and technical knowledge of scientific ballooning. His knowledge has developed through "a long continuous interest and participation in the field of scientific ballooning and interaction with the many present and past commercial and government organizations" [3]. The knowledge and experience of individuals within a specialized field is what sustains the use and value of grey literature within the field. Jim Winker's large-scale interaction with balloon literature clearly exemplifies this.

Mr. Winker visited over 60 sites in his effort to find and select documents for the library [4]. He started locally with his own collection, the Raven Industry library, private collections in Sioux Falls, South Dakota and the holdings of Wallops. His search expanded to include the information repositories of the National Technical Information Services (NTIS), Defense Documentation Center (DDC), Association of Balloon and Airship Constructors (ABAC), and American Institute of Aeronautics and Astronautics (AIAA).

He explored organizations that use balloons currently or have used them in the past. These organizations included research centers, manufactures, scientists, research sponsors, and operations centers. In his effort to be comprehensive, his search included an "everything else" category as well. He evaluated the resources of libraries, museums, individuals, published works, and international resources [1].

With the care of an archivist, Mr. Winker researched preservation quality paper, durable binding, and copyright implications [5]. While he realized the selection and preservation of the documents was of fundamental importance, he also realized that "the database may well be the most important part of the project. Without it, the library would have a far more limited benefit" [1]. The database would promote the existence of the collection and ultimately provide access to the documents. Working with Janet Ormes, GSFC Library Head, Mr. Winker, selected software to support the database and created guidelines for cataloging the documents. The *COSATI Standard for Descriptive Cataloging of Government Scientific and Technical Reports* was relied upon for the creation of the rules of entry [6]. For each document, an entry was created for the database using a variety of descriptors intended to aid in the process of information discovery.

The askSam database was chosen in 1994 and began production with the database successfully organized and defined [7]. At this same time, database search methods were explored and decided upon. Progressive for the year, both controlled vocabulary and natural language searching were made possible [8]. In the ten years that the documents were collected and that the database grew in significance for the ballooning science community, technology was developing that would further enhance the work of Mr. Winker.

Current Status of the Documents

Since the Wallops' Balloon Technology Library began, the Goddard Library has become increasingly involved in developing digital archiving capabilities. The Library developed the Digital Archiving System (DAS), a prototypical infrastructure for creating a combined metadata repository allowing for heterogeneous digital objects to be searched with a single search mechanism and presented in a single results page. The DAS is an institutional repository of scientific and technical information including selected Goddard web sites, videos, images and documents. With this, the opportunity has been presented to expand the usability and accessibility of the balloon technology library.

The documents exist primarily in paper format and are organized according to the accession number. The documents are housed as a separate collection at the Wallops Island Technical Library. The database is accessible to all from the library's website. NASA researchers can gain access to the documents by either using the library at Wallops or having the documents scanned and emailed to them. Researchers outside of NASA can make requests from Wallops with some restrictions placed on their access.

Since January of 2004, some balloon technology documents have been scanned as a pilot project into a MySQL database as PDF files. The askSam records as contained in a WAIS database link to the PDF files; however, the DAS which the MySQL database supports does not relate to the askSam database. The askSam software is no longer being developed. It has limited capabilities for interacting with the World Wide Web and the searching environment. Currently we have two separate search interfaces, one is a meta search in WAIS and the other Autonomy which indexes the documents and has full text search capabilities [9,10]. WAIS and Autonomy can't be configured to work together to provide meta and full text searching. The DAS can do both by using MySQL and Lucene. Not to mention, the DAS allows you to search, archive, and preserve documents for future use. By incorporating the balloon technology documents into the DAS, the balloon science community will have increased visibility and access through meta, full-text searching and retrieval options to include the full-text.

The project is currently at a crossroads where decisions will have to be made regarding the incorporation of the database into the DAS and the accessibility of the documents on many different levels and perspectives. The authors of this paper turned to the community involved in the creation, the delivery and the use of the library and database to determine the future direction of the project.

Knowledge Management Assessment

An assessment survey based on the Knowledge Management Assessment Tool (KMAT) from the American Productivity & Quality Center was administered to nine individuals, representing different perspectives on the project [11]. Further, these same individuals assessed the project based on the recommendations that The Scholarly Publishing & Academic Resources Coalition (SPARC) put forth in "The Case for Institutional Repositories: A SPARC Position Paper [12]. See Appendix A for the complete survey assessment. A meeting was held on August 12, 2004 with survey participants and Balloon Program Office (BPO) representatives in attendance. Issues brought forth by the assessment created the agenda for the meeting.

Given that the work of the Balloon Technology collection is a small slice of a greater organizational structure, survey participants were ask to be cognizant of this while responding to the KMAT as knowledge management is the function of an organization. The authors felt that the current

status of the project presented an instance of the process of knowledge management in play. Mr. Winker is retiring for the second time in January 2005 and will no longer be working on the project. As stated earlier, he has performed the vast majority of document collection and cataloging. The authors wanted to examine how well the knowledge that Mr. Winker possesses is being transferred to the individuals remaining on the project.

The greatest benefit of the knowledge management assessment indicated an important difference in perspective among survey participants. Two perspectives emerged: content and infrastructure. Individuals with a content perspective included contributors and users of the information contained within the collection. Individuals with an infrastructure perspective included systems, web, library, and archival staff that primarily work to make the collection accessible to users of the collection.

At the assessment meeting, both perspectives were defined and the importance of each established. Certain points of confusion regarding current communication between all project participants were made clear. Traditionally, the individual who creates information and/or collects the information is not the same individual that catalogs, stores, and distributes the information. For ten years, Mr. Winker was performing most tasks related to the collecting, cataloging, and searching of the documents, not to mention creating literature as well. Enter the possibility of the DAS and open access archives and the flow of knowledge from old processes to new processes does not transfer without some adjustment. The second part of the assessment devised from the SPARC paper complimented the first as it helped the discussion participants describe the processes under consideration for change.

Can't see the Repository for the Documents

For the purpose of this paper, the definition of an institutional repository as put forth by the Scholarly Publishing & Academic Resources Coalition (SPARC) will be used. Institutional repositories are digital collections that capture and preserve the intellectual output of a particular community. Repositories are intended to expand access and retain control over the scholarship produced by that community. Further, the repository has the potential to contribute to the community or the institution it is a part of by providing tangible indicators of the community's quality through the demonstration of the scientific value of its research activities [12].

Being a part of the DAS can in and of itself increase the visibility, status and public value of the balloon science community. Money, time, and technology aside, the documents themselves create the biggest obstacle to creating an institutional repository at Goddard as put forth by SPARC. The SPARC paper further defines an institutional repository as "open and interoperable." To promote interoperability and open access, institutional repositories provide access to the broader research community through no or low barrier access. Either by providing a search mechanism with indexing or by maintaining and exposing metadata to be harvested by other institutions, interoperability is gained among institutions [12].

While the BPO is interested in sharing the database of records in the manner described above, the office is not interested in sharing the documents in a global capacity. The proprietary nature of some of the documents and the competitive environment in which they were created prohibit the collection of documents as a whole from being accessed outside NASA IP ranges. Small Business Innovation Research (SBIR) documents are an example as they are restricted for five years after their release. The BPO representatives want to keep the collection primarily for balloon scientist with a relationship to NASA research, hence limiting access to greater research community. It is extremely likely that other Goddard projects will have similar concerns or restrictions regarding their documents. Fortunately, the technology exists to allow different levels of access to different types of information allowing Goddard to interoperate fully in at the metadata level.

If the database records are to be open and interoperable, the existing data fields from askSam will have to be mapped to the metadata elements of the Goddard Core. The Goddard Core is a metadata element set and is employed in the DAS single search mechanism. Metadata is information that describes a digital object; like a library catalog card for a digital object. The Goddard Core Metadata set contains 24 elements that describe project-related objects of interest to GSFC. It extends ISO 15836 (Dublin Core) Standard for Descriptive Metadata for Electronic Resources while using Open Archives Initiative standard protocols that can interoperate with other systems.

When Mr. Winker was devising his catalog entries, he was working in a contained environment. There was no need to consider how the descriptive fields he employed related to a larger information structure. The technology did not exist for interoperability to be an issue. The Goddard Core establishes guidelines for organizing documents and other objects in a way that is both meaningful and accessible across GSFC, within NASA and in the open access environment.

The mapping will involve more than a programming solution as the fields used in askSam do not match directly or in some cases not at all to the Goddard Core elements. Without ISO standards in place, definitions of descriptors were not nearly as controlled. The Goddard Core allows for project profiles so if need be the 24 elements can be expanded to accommodate the particularities of a document or object collection. The mapping of askSam fields to Goddard Core is a key area for the collaboration from both content and infrastructure perspectives.

Copyright presents another problem in bringing an already existing print library and database into the realm of the online repository. To date, there are 4,330 records in the database of which 847 are indicated as copyrighted. That is roughly 20%. A copyright campaign will need to be launched on behalf of the repository under the auspice of the BPO. This will take a significant amount of time and agreement is not guaranteed. It was agreed that copyrighted documents will begin to populate the DAS and as permissions are gained documents will be added with the understanding that all documents are not likely to be included in the online repository.

As mentioned previously, Mr. Winker is retiring. His knowledge is the current certification and collection process. While encouraged by the ease of collecting new submissions to the collection through online technologies, the users of the collection and the BPO representatives were concerned about the quality that self-selection produces as well as the possibility for submitting items out of scope. Within the balloon science community, there is considerable interest in expanding the collection to include literature related to the experiments conducted using the balloons. Further, online submissions would not replace the need to produce an archival paper copy for the balloon technology library. The meeting participants decided that an official joint committee of Goddard librarians and BPO staff should be sanctioned to establish a certification process and to address any further decisions that need be made regarding the repository. Additionally, the joint committee will facilitate and encourage communication between infrastructure and content perspectives.

Conclusion

As the survey participants sat in the room where the Balloon Technology library is housed, a collective realization fell over the group at the conclusion of our assessment discussion. The ways in which people seek and interact with information are dynamic and online. The print collection resting on the shelves seemed anachronistic given our discussion. The BPO and the Goddard Library agreed to move forward with scanning the documents, mapping the records, seeking copyright permissions, and forming a joint committee. All have agreed on the importance of digitizing this collection to the balloon science community. Further, it was agreed that once complete, the addition of the balloon documents to the DAS institutional repository could serve as a model for other NASA and/or government projects trying to organize, collect and preserve specialized knowledge that manifests largely in grey literature.

Currently, funds are being pursued to make the transition from a library and a database to an institutional repository a reality at Goddard. Given the nature of the types of information produced at Goddard, any institutional repository created at Goddard or a like environment will not be able to participate fully in open archive initiatives. In our case, the DAS works as an institutional repository for Goddard and to a lesser degree beyond the walls of NASA IP addresses: metadata is acceptable for web harvesting but most documents would not be freely available.

The DAS is still in beta testing and contains over 90,700 web pages, 900 images, and 400 videos. The Balloon Technology document collection will be the first document collection to be included in the DAS. The choice of this collection has proven fortunate in that it exposed many of the issues we would likely encounter when considering the addition of other collections.

Author Information

Ms. Andrea C Japzon
Access & Navigation Services Team Leader
Information International Associates, Inc.
NASA Goddard Space Flight Center
Code 292, Bldg 21, Rm L112
Greenbelt, MD 20771
USA

ajc1170@aol.com
<http://www.gsfc.nasa.gov>

Works Cited

- (1) Winker, J. A. (1994). A Balloon Technology Library Database. Paper from the 32nd *Aerospace Sciences Meeting & Exhibit*. Reno, NV, January 10-13, 1994.
- (2) Winker, J. A. (1993). Balloon Technology Library Project Monthly Progress Report No. 1. December 3, 1993.
- (3) Winker, J. A. (1993). Balloon Technology Library: Discussion topics for initial project meeting of 23 November 1993. November 18, 1993.
- (4) Winker, J. A. (1998). Balloon Technology Library Project Monthly Progress Report No. 45. January 2, 1998.
- (5) Winker, J. A. (1994). Balloon Technology Library Project Monthly Progress Report No. 3. February 1, 1994.
- (6) Suggested Guidelines. (1994). Condensed from *Guidelines for Descriptive Cataloging of Reports: a Revision of COSATI Standard for Descriptive Cataloging of Government Scientific and Technical Reports*.
- (7) Winker, J. A. (1994). Balloon Technology Library Project Monthly Progress Report No. 5. April 2, 1994.
- (8) Winker, J. A. (1994). Balloon Technology Library Project Monthly Progress Report No. 6. May 4, 1994.
- (9) *Balloon Technology Database*. (2004).
<<http://library.gsfc.nasa.gov/Databases/Balloon/balloon1.html>> (October 26, 2004).
- (10) Search Goddard Library. (2004)
<<http://library.gsfc.nasa.gov/SiteSearch/premium/premiumquery.html>> (October 26, 2004).
- (11) American Productivity & Quality Center. (2001) *The Knowledge Management Assessment Tool (KMAT)*. <http://www.kwork.org/White_Papers/KMAT_BOK_DOC.pdf> (October 26, 2004).
- (12) Crow, Raym. (2002). The Case for Institutional Repositories: A SPARC Position Paper. <<http://www.arl.org/sparc/IR/ir.html>> (October 26, 2004).

Appendix A

Part I

The KMAT was used as presented at http://www.kwork.org/White_Papers/KMAT_BOK_DOC.pdf with the following questions removed: P3, L2, M1, and M3. Reason being the project is produced by a government organization and therefore will not be sold or marketed for profit.

Part II

1. What purpose does an Institutional Repository serve?
2. Policy consideration:
 - a. What is the current copyright policy? Does it need to be improved?
 - b. How are restricted documents handled? Could this be done differently?
 - c. With regards to accessibility, do we need a policy that differentiates between internal and external customers?
 - d. What is the current certification process, the process that assures the quality of the documents added to the repository? How will this change once Mr. Winker leaves the project?
 - e. When you consider the future of the repository, will we need a formalized accession policy?
3. Does the repository embody the institutional quality of
 - a. The Balloon Technology Program YES ☐ NO ☐
 - b. NASA YES ☐ NO ☐
 - c. If No to either "a" or "b", please explain what needs to change for this to be so?
4. Does the repository have formal or official recognition in the Balloon Technology Community? Please explain why or why not.
5. On a scale of 1 to 5, please indicate how important you think it is that the Balloon Document Repository reflects the following. 1 being not at all important and 5 being very important.
 - a. Institutionally Offered _____
 - b. Scholarly _____
 - c. Cumulative and Perpetual _____
 - d. Open & Interoperable _____
6. Please indicate by checking which of the following you think should be reflected in the metadata? Jim Winker's Metadata Elements have been mapped to the corresponding Element in the Goddard Core.

Jim Winker's Metadata Elements Used	Goddard Core Metadata Elements Used	Please Check X
Access Number	Identifier.Persistent (Auto)	
Title	Title	
Author	Creator.Employee	
Responsible Organization	Creator.Organization	
Funding Organization	Contributor.Organization	
	Creator.RecordCreator (Default)	
Date	Date	
Notes		
	Date.Current	
	Date.Created (Auto)	
	Date.RecordCreated (Auto)	
Report Number	ID	
Contract Number	Contributor.Contract	
Description		
Subject Terms	Subject.Uncontrolled	
Content	Description	
Available From		
	Contributor.Code	
	Type (Default)	
	Content.Type	
	Format (Auto)	
	Subject.Discipline	
	Identifier.URL (Auto)	
	Language (Default)	
	Rights (Default)	
	Audience (Default)	

7. Please describe any other elements that you think should be added?
8. Please feel free to make any additional comments regarding the Balloon Technology Documents?

Making Grey Literature Available through Institutional Repositories

Nathan Rupp and Lee LaFleur
Albert R. Mann Library; Cornell University

Grey Literature at Cornell

The Cornell University Library exists to support the three-fold mission of the University: research, education, and extension (Cornell is the land grand institution for the State of New York). It supports this mission through the services it provides and the material it collects; it collects material produced by the university itself as well as outside publishers. In recent years, university librarians have recognized that a significant amount of material worth collecting in support of the university's activities can be classified as "grey literature." Although grey literature has in the past been considered too ephemeral and marginal to spend time collecting, it is part of a communications process that, although separate from mainstream publishing, is just as important. Even though some information resources are created with nontraditional publishing methods does not mean that they are unimportant; often, these grey literature resources later move "into the mainstream of information products." (1) In addition to its general value in research, grey literature is particularly important in the sciences. "The quest for scientific knowledge is an evolutionary process in which every increment of new knowledge adds to, modifies, refines, or refutes earlier findings," and grey literature is an important part of this process. (2) To this end, librarians at Cornell, including librarians supporting the sciences, have determined that collecting grey literature is as important to supporting the university's activities as is collecting any other resource.

A recently published Cornell University Library report surveyed the types of grey literature currently being produced within the university. These included the physics pre-prints contained in the arXiv e-print service; documents created by Cornell University Cooperative Extension; working papers produced by programs such as the Cornell Food and Nutrition Policy Program and the Center for Advanced Human Resource Studies; and proceedings produced in association with a number of conferences sponsored by Cornell.

Conference proceedings are an important type of grey literature for a number of reasons. Conferences and the proceedings that document their programs are a valuable part of the scholarly communications cycle. The proceedings from conferences document the current state of research and provide a lasting record about what took place at a particular conference. Furthermore, proceedings enable research to be shared among those who were not present at a conference and serve as a point of reference for those who were in attendance. Many research findings presented at conferences and recorded in conference proceedings are not reported anywhere else. However, conference proceedings can be difficult for researchers and information professionals to identify, locate and acquire. Many times the groups that organize conferences are small and their events are not widely publicized. Even those proceedings collected by libraries can often be difficult to access due to lack of item level indexing, frequent title changes and the irregular and informal nature of their publication.

Conference Proceedings Project

As the report surveying grey literature at Cornell was being researched, other library staff members had begun to think about how the library could provide online access to proceedings published in association with conferences held at Cornell. They designed a pilot project in which access to Cornell-produced conference proceedings could be made available online, and in doing so, identified a number of activities such a project would need to undertake. These included the identification of a set of conference proceedings with which to establish the project and the identification of a system with which to make the proceedings available online. Although conference proceedings are being published across a number of colleges at Cornell, the project focused on surveying conference proceedings produced by the Life Sciences libraries. To identify a small set of conference proceedings with which to begin the project, a number of Life Sciences librarians and faculty were interviewed. In the end, the proceedings for the Wine Industry Workshop, a wine production conference held annually in Geneva, New York, were chosen for the project.

DSpace

Once the set of proceedings was chosen, a survey of the various digital library platforms available to provide online access to the proceedings was begun. This examination showed that Cornell was experimenting with a wide variety of digital library platforms. These included DLXS, developed by the University of Michigan and used by Cornell to support its Core Historical Literature of Agriculture and Home Economics Archive projects; DPubS, an open source digital library platform developed at Cornell and used to support Project Euclid; FEDORA, developed by the Digital Library Research Group in the Cornell Department of Computer and Information Science in partnership with the University of Virginia;

and DSpace, the institutional repository developed by MIT and Hewlett-Packard and put into experimental use at Cornell. (3,4,5,6)

DSpace was chosen for the project for a number of reasons. First, Cornell University Library had recently installed an instance of DSpace and was looking for some projects to experiment with. Second, one of the members of the committee awarding the internal grant that funded this project was an administrator in CUL's IT department and she recommended DSpace. Third, DSpace's hierarchical structure of "communities" and "collections" was viewed as useful for structuring the proceedings. For the project, the proceedings were to be structured in such a way that researchers could search the collection from numerous levels:

- Entire collection of proceedings;
- Conference proceedings produced in one particular college (Agriculture and Life Sciences, for example);
- Proceedings from one *particular* conference in a single college;
- Single year's set of conference proceedings from a particular conference; or
- Individual proceedings within a single year.

(cf. Diagram One)

Fourth, DSpace's metadata creation tool seemed very useful in assisting those adding proceedings to the repository. In the past, metadata librarians at Cornell have had the opportunity to work on digital library projects in which authors of digital information resources also created the metadata associated with them. The librarians have learned that sometimes the authors create less than perfect metadata which subsequently needs to be cleaned up and massaged for inclusion in the digital library system. Any tool that enables easier, more efficient creation of metadata is useful.

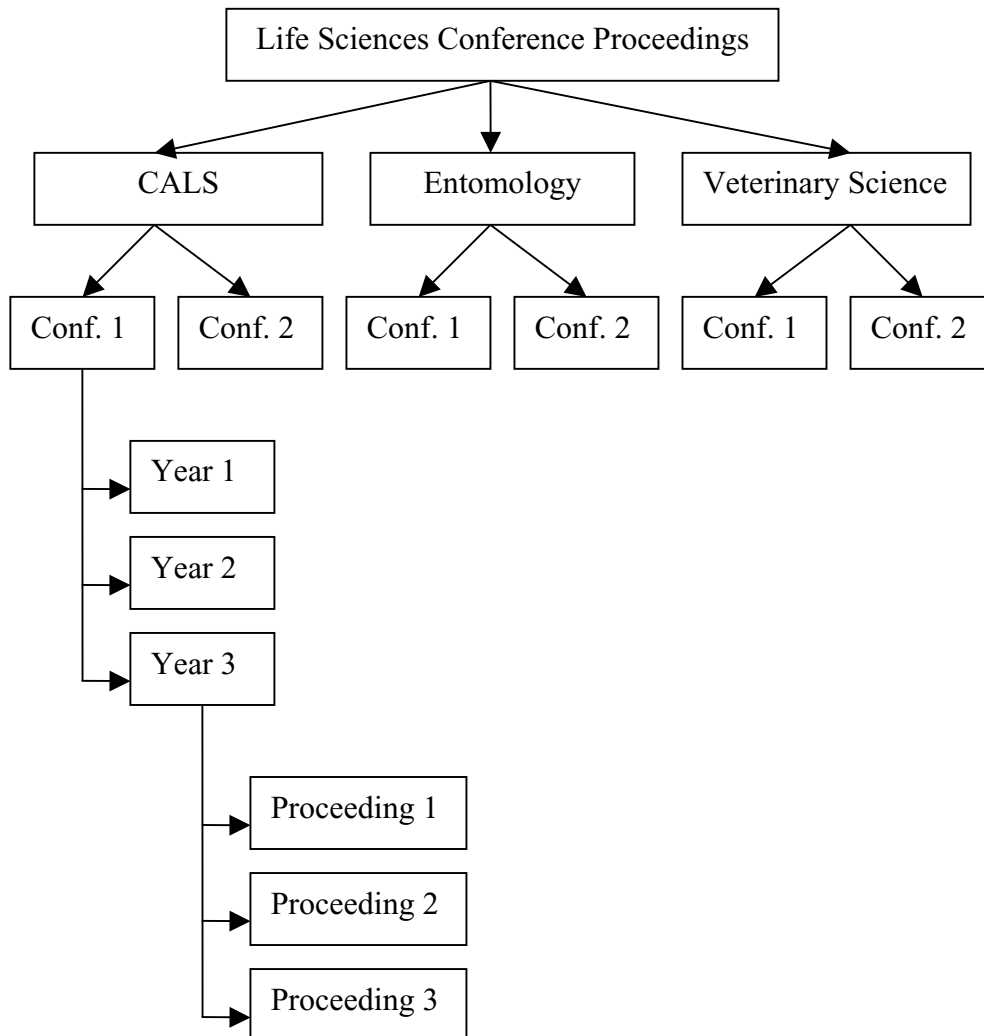


Diagram One: Hierarchical Nature of Conference Proceedings.

Gathering and Scanning Proceedings

After a set of conference proceedings and an institutional repository in which to store them were selected, the proceedings themselves were collected. This process involved contacting the administrative staff and the faculty member responsible for the conference; the staff was able to provide the library with both print and digital copies of four years of proceedings. This ad-hoc method of gathering the proceedings worked well for this project, but if the project were to be made a permanent part of the library's programs, a specific workflow would have to be developed for the identification and gathering of proceedings. The program managers would have to work with the library director of public relations and engage in one-to-one marketing to create awareness on campus of the conference proceedings repository. Methods would have to be designed for moving the proceedings from the authors' desktops to those of the librarians administering the project. A system would need to be developed for storing the proceedings in paper or digital format between the time of their receipt and the time they are actually loaded (in the case of digital files) or scanned for loading (in the case of print copies) into the repository. Scanning proceedings for loading into a repository is not inconsequential; the project managers would have to select a scanning vendor from the vendors available in-house at Cornell and outside the University. Care would have to be taken to ensure that the scanned files received for inclusion in the library were accurate and of good quality. Fortunately, Cornell University Library in general—and Mann Library in particular—has a great deal of experience working with scanning vendors.

Loading Proceedings into DSpace

After the proceedings were collected, they were loaded into DSpace. Through this process, it quickly became evident that DSpace may not have been the best digital library platform/institutional repository to use for the project. First, DSpace's hierarchy proved not to be extensive enough for the purposes of the project. The hierarchy in the initial release of DSpace was limited to just two levels, "communities" and "collections." Although the hierarchy was expanded to three levels with the addition of "sub-communities" in a 2004 release, neither hierarchy was extensive enough for the purposes of this project. At least four levels of hierarchy would be needed to provide useful access to conference proceedings in the life sciences project. While DSpace is based on the Dublin Core metadata scheme, which incorporates *hasPart/isPartOf* refinements to its "relation" element, the use of these refinements to provide access to the various levels of the hierarchy of materials is not the best solution to this problem. DSpace would be more useful if it could be modified to support hierarchies of different sizes depending on the collection in question.

In addition, DSpace's metadata creation form proved to be less useful than was originally thought. It was lengthy and complex; even the metadata librarian working on the project found that using the form to create metadata for the proceedings was tedious. Authors of conference proceedings not skilled or interested in metadata creation would find this form even more of a barrier. It would be more practical to design a metadata creation form that still utilized Dublin Core but was better suited to the needs of local projects. Such a form could be used to create a single metadata file representing metadata for all the objects in a single repository; this file, in turn, could be batch loaded directly into the DSpace repository, bypassing the DSpace metadata creation tool.

On the other hand, it is possible that *no* metadata creation form could be designed to enable proceedings authors to easily create metadata for their proceedings. A number of current digital library projects that receive content from external authors have needed to assign a project staff member the task of editing the metadata accompanying any content received. This may also need to be done for the proceedings project: a librarian may need to coordinate the creation of metadata for the proceedings upon their receipt. A tool best suited to meet the metadata creation needs of *library staff*, rather than proceedings authors, may need to be developed. The characteristics and features of a metadata creation tool used by library staff will differ from those of a tool intended for use by proceedings authors external to the library. The process in which metadata is to be created—whether it is to be created by the authors of the proceedings or by librarians after the proceedings are submitted—will need to be determined before any metadata creation tool is developed. The development of a metadata creation tool could mean a refinement to the DSpace tool or the development of an entirely different one.

In this study of DSpace, it was learned that its preservation mechanism (7) —converting the objects within the repository into bitstreams and preserving the bitstreams—has been called into question by a number of preservation professionals. Even though bitstreams may be easier to preserve for long periods of time than are, for example, Microsoft Office documents, the process whereby the objects within the repository are converted to bitstreams and then converted back into objects needs to be carefully evaluated. It is one thing to convert a Microsoft Office document into a bitstream for preservation and storage in DSpace and then convert that bitstream back into a Microsoft Office document. It may be something else entirely to convert a Microsoft Office document into a bitstream for preservation and storage in DSpace and then, in the future, convert that bitstream into a format that *does not currently exist*.

As consideration has been given to the design of a library service based on the findings of the proceedings project, DSpace's limitations have prompted further exploration of other digital library systems. There are advantages and disadvantages to each of the digital library/ institutional repository

options currently available at Cornell. The DSpace designers have gotten the hierarchical design of the system right; they just need to expand it. Another repository that has a built-in hierarchy and is structured to work with journals is DPubS. Since DPubS is an open source platform, it will improve with time as implementers add new features. In addition, DPubS is already in production mode, so those using the DPubS platform for other projects will be able to rely on the expertise of those currently working with the system. Even though DLXS is also supporting projects that are in production, it does not have a DSpace-like hierarchy built in and seems to work best with monographic materials. FEDORA is an excellent digital library platform, but it only supports a repository back end; the development of a user interface for any FEDORA-based system would involve a major commitment of time and money.

Economic Concerns

In addition to the financial investment required for the development of a FEDORA-based system, similar investments may be required for a number of other pieces of this project. Scanning back runs of print proceedings for inclusion in the repository would not only involve the selection of a vendor but also the design of a business plan to pay for the scanning of the proceedings; scanning costs vary significantly among vendors. In addition, if a library service was developed in response to this project, decisions would have to be made about how these scanning costs would be recovered. The costs could be passed on to the authors of the proceedings or even users of the site. Another cost that would have to be addressed is the cost of proceedings for *current* conferences. When the repository was demonstrated for an audience at the 2004 Wine Industry Workshop, the question most frequently asked was whether the proceedings should be made available for a fee. Workshop attendees currently pay for the proceedings as part of the conference registration fee. If the proceedings were made freely available via an institutional repository to both workshop attendees and non-attendees alike, the non-attendees would have an unfair advantage, since they would not have already paid for the proceedings through the conference registration fee. On the other hand, one could ask if a charge would still be levied for the proceedings if they were made available online. If that charge is currently being levied only to cover the production and binding costs, then that charge could be eliminated if the proceedings were made available online and not produced in paper form and made available at the conference. The costs of "producing" the proceedings would then fall to the conference attendees, who would need to go to the repository, download the proceedings, and print them out themselves.

Additional Technical Issues

If the system needed to distinguish between those who had attended the conference and had "free" access to the proceedings and those who had not attended and did not have access, different levels of access would need to be built into the system. This might be accomplished with user recognition devices such as user names and passwords, which may not be built into a particular digital library/institutional library system (it is not built into DSpace). This would be an important design consideration in the development of any system.

A second technical consideration involves the life cycle of conference proceedings. Abstracts for conference papers are sent to the conference conveners before the conference, Microsoft PowerPoint slides are often created in connection with the papers, and the actual papers are submitted to the conference conveners separately. Each conference "proceeding" may consist of these parts; if a conference proceedings repository recorded these separate parts, there would need to be a way to tie them together so a user searching for a proceeding would be able to retrieve all components of the proceeding. Similarly, the repository would have to be structured in such a way that it could store and provide access to various types of content such as Microsoft Word documents, Microsoft PowerPoint slides, and other types of files.

A third technical consideration involves the extent of any collection of conference proceedings. It is one thing to create a repository for a single institution's proceedings, but a much larger effort to create a network of repositories storing proceedings from multiple institutions. This would be a logical extension of any proceedings repository; no single institution sponsors all conference proceedings, so it would be beneficial to network a number of proceedings repositories together to provide shared access to a large number of conference proceedings. Institutions would need to partner with one another to create a shared repository of conference proceedings. In addition, the conference proceedings repositories would need to interface with one another; this would be another important design consideration in the design of a conference proceedings repository. Standards and protocols like the Open Archives Initiative Protocol for Metadata Harvesting would need to be implemented in each one of the repository systems to make them compatible with one another.

Conclusion

In conclusion, to fully support the three-fold mission of Cornell University, the library needs to provide access to *all* types of information resources, including grey literature that is not as readily available as other information. One particular type of grey literature that has been identified as difficult to access is conference proceedings. This project explored the use of a particular institutional repository—DSpace—in providing access to these conference proceedings. This project enabled Cornell University librarians to learn about the issues surrounding the use of an institutional repository—or other digital library platform—to provide access to conference proceedings. This project's findings will be used to determine whether or not a service of providing access to conference proceedings via a digital library platform is viable within the Library.

Notes

1. Dankert, Philip; Gregory Lawrence; Patricia Viele (2003). *Grey Literature: A Research Report (2002-2003 CUL Internal Grant)*. Ithaca, N.Y.: Cornell University Library, p. 1.
2. Weintraub, Irwin (2001). The Role of Grey Literature in the Sciences.
<http://library.brooklyn.cuny.edu/access/greyliter.htm/>
3. DLXS: Digital Library Extension Service. <http://www.dlxs.org/>
4. DPubS. <http://dpubs.org/>
5. The Fedora Project: An Open-Source Digital Repository Management System.
<http://www.fedora.info/>
6. DSpace. <http://www.dspace.org/>
7. Bass, Michael J, et al. (2002). *DSpace – A Sustainable Solution for Institutional Digital Asset Services – Spanning the Information Asset Value Chain: Ingest, Manage, Preserve, Disseminate: Internal Reference Specification: Functionality*. Cambridge: Hewlett Packard Company; Massachusetts Institute of Technology, p. 4-5.

Author Information

Mr. Nathan Rupp
Metadata Librarian
Albert R. Mann Library; Cornell University
Ithaca, NY 14853
USA

nar25@cornell.edu
www.cornell.edu

Building an autonomous citation index for grey literature: the Economics working papers case

José Manuel Barrueco, Universitat de València, Spain
Thomas Krichel, Long Island University, New York, USA

Abstract

This paper describes an autonomous citation index named CitEc that has been developed by the authors. The system has been tested using a particular type of grey literature: working papers available in the RePEc (Research Papers in Economics) digital library. Both its architecture and performance are analysed in order to determine if the system has the quality required to be used for information retrieval and for the extraction of bibliometric indicators.

1.- Introduction

The main characteristic that differentiates the scientific literature from other literary representations is the relationship between documents established through citations and bibliographic references. The scholarly work can't exist on its own. It must always be related to documents in the same subject area that have been published earlier on. In this way we can see the literary corpus as a complex semantic network. In that network, the vertices are documents and the edges are citations and references.

It is important to differentiate between citations and references. Citations are referrals that a scientific work receives from other documents published later on. References are referrals that one document makes to other works published before.

In the 1960s Eugene Garfield developed the first tool devoted to the representation of relationships between scientific documents: the Science Citation Index. Since then, citation indexes have become an important study tool in some areas. In Scientometrics, citation indexes have become an essential tool for the evaluation of scientific activity. In Information Science researchers have studied the possibility of browsing the scientific literature using references and citations. In this way, once an interesting document has been found, it would be possible to use its references to find similar ones.

Compiling large scale citation indexes for printed literature, using human labour, has been an expensive task. In the past only the ISI (Institute for Scientific Information) has carried out this type of work. However, nowadays all scientific documents are generated in electronic form. If they are available on the Internet this allows the possibility of extracting the references automatically. The references of a scientific paper identify the cited documents and create the appropriate links if they are available in electronic format. With such system the costs would be dramatically reduced and new indexes covering new document types (i. e. grey literature) could arise.

The pioneers in this research area were Steven Lawrence and C. Lee Giles with the CiteSeer autonomous citation index (ACI) for Computer Science. They define an ACI as a system which "can automatically create a citation index from literature in electronic format. Such a system can autonomously locate articles, extract citations, identify citations to the same article that occur in different formats, and identify the context of citations in the body of articles. The viability of ACI depends on the ability to perform these functions accurately". In this paper we describe a similar system called Citations in Economics (CitEc). This system uses CiteSeer technology to automatically build a citation index for documents contained in the RePEc (Research Papers in Economics) digital library.

The remainder of this paper is organised as follows. Section two describes the RePEc data set which has been used as test bed for the citation index that we have developed. Section three describes the CitEc architecture. Section four is devoted to the analysis of the system performance in order to determine whether it could be used to extract bibliometric indicators. Otherwise it would be limited to information retrieval. Section five concludes the paper.

2.- RePEc: a digital library for Economics

RePEc (Research Papers in Economics) is the largest decentralised non-commercial academic digital library in the world. Its home page lives at <http://repec.org>. A gateway that is compatible with the OAI-PMH (Open Archives Initiative, Protocol for Metadata Harvesting) (<http://www.openarchives.org>) is available at <http://oai.repec.openlib.org>. RePEc describes two types of documents: grey literature, namely working papers, and articles published in peer-reviewed journals. In November 2004 there were 140.000 working papers and 144.000 articles. RePEc is based on a distributed architecture where research institutions worldwide share information about the documents they publish. The details are contained in two documents: the Guilford Protocol and ReDIF (Research Documents Information Format).

The Guilford Protocol (GP), named after the town where it was created, is a set of technical requirements that an institution should accomplish to become a member of the RePEc community. It covers only institutional collaboration, i.e. individuals can not join RePEc. There are two ways for an institution to participate in RePEc: archives or services. Archives collaborate by providing bibliographic information about the documents their institution publishes. They also may provide the full-text of these papers. Technically an archive is a space in the hard disk reachable by an HTTP or FTP server. There, files containing bibliographic information are stored. The structure of this space is defined in the GP.

The second pillar of RePEc is ReDIF. All data transmitted between archives and services is encoded in a specific bibliographic format named ReDIF. ReDIF was created to meet the RePEc needs and therefore it is not aimed to become a widely used format for interchange of bibliographic data between libraries. Its main characteristic is that it is simple enough to be used by non-technical people outside of the library world. This is because the most archives are maintained by administrative staff or academics without knowledge of library procedures.

ReDIF allows to describe working papers, articles in journals, physical persons and software components. Each object is described by a template made up of several fields like a traditional database record. They use an attribute: value syntax. Fields can be optional or mandatory. The main mandatory field is the Handle which holds a code that identifies the object within the RePEc dataset.

RePEc was created in 1997 from a collaboration of several projects working on electronic document dissemination in the discipline. Since then the number of institutions collaborating as archives has been increasing. At the time of writing, in November 2004, there are 413 archives. The number of documents by archive depends on the kind of institution it belongs to. For example, the prestigious NBER (National Bureau of Economic Research, USA) provides a very large archive describing all the 10897 papers it has published.

Metadata as it appears in the archives is of little utility for researchers. Further processing is needed to take the information and present it in a user-friendly way. This is the objective of the user services. User services are the main way a user works with RePEc. A complete list of user services can be found at the RePEc home page. They add value to the data provided by the archives in several ways:

- Scanning for new data and creating an awareness system to announce new additions.
- Creating a searchable database of papers.
- Creating some type of filtering service to help the user in the selection of the most relevant documents.

Building a citation index allows to create an additional user service that has citations as its prime focus. This is done in the CitEc project. Its home page can be found at: <http://netec.ier.hit-ac.jp/CitEc>.

3.- Citations in Economics architecture

The CitEc architecture is based in two main elements as it is shown in Figure 1. First, we have a knowledge base¹ where all authoritative metadata about RePEc documents is stored. This base represents the main improvement we have implemented in the CiteSeer software. The quality of the bibliographic references provided in the papers is variable. For instance, it is usual to find different forms for the same author names, journal titles, etc. We use the knowledge base to complete and improve the references quality with metadata provided by the publishing institutions. Secondly, we have a series of three software modules, one for each step in the reference linking process:

- Collecting metadata and documents' full text.
- Parsing ongoing documents in order to find the references section, to identify each reference and to extract their elements (authors, title, etc.).
- Linking of references with the document full text they represent if available on RePEc2.

It is important to note that each module is based on the output of the previous one. In this way, the successful processing of each document implies to successfully surpass the sequence of three levels.

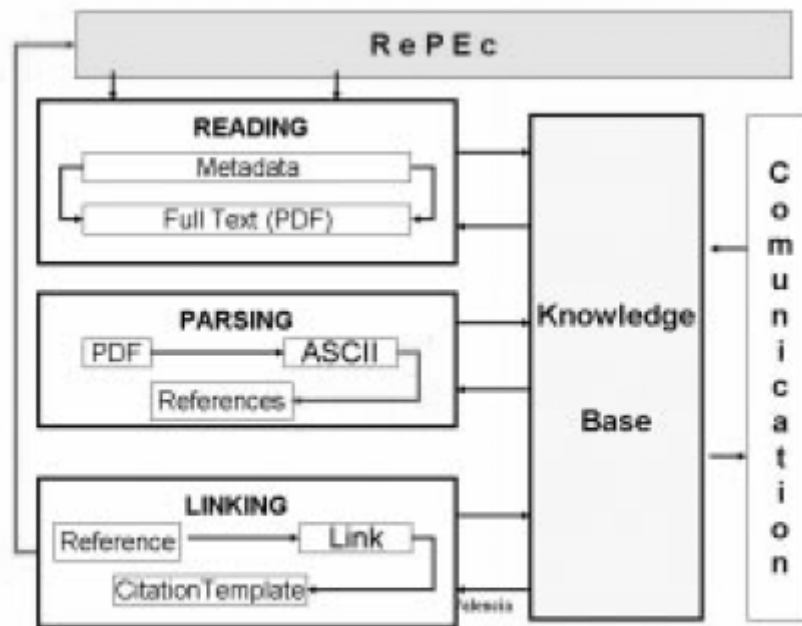


Figure 1: CitEc Architecture

3.1.- Collecting

Collecting involves three different steps: (1) to collect the documents' metadata, (2) to download the documents' full text and (3) to convert them to a parseable form.

The metadata quality varies from archives. There are archives that provide very complete records for each paper, including abstracts, JEL (Journal of Economics Literature Classification) codes, etc. On the other hand, other archives may only provide titles and authors. There are three main problems with the metadata that seriously affects the processing of papers:

- The absence of publication dates. This field is optional in ReDIF and some archive maintainers don't use it. In our research this data is fundamental because the publication year is one of the attributes we use to check whether a citation goes to a RePEc document or not. Fortunately the publication year usually forms part of the working paper number. Most series are numbered like: 9901, 9902 ... Taking advantage of this convention, we have developed procedures that guess the year from the paper numbers.
- The format in which the author names are written. ReDIF requires that each name be placed in a different field but some archive maintainers write all authors in a single field, separated by some punctuation. We have also developed procedures to cope with such problems and to correct them as far as possible.
- Wrong URLs. The URLs provided by the archives to retrieve the documents full text are incorrect. This is a rare but serious problem. If we can not access the paper is not possible to circumvent the problem as we have done with the other cases.

We are working with a distributed library of metadata. There is no a single place where all full text documents live. They are dispersed in multiple servers from multiple institutions. Therefore the second step is to download to our hard disk those documents that are available in full text. This is done by going trough each archive, reading the bibliographic information and, if a File-URL field is found, retrieving the resource contained in the URL. Usually such resource will be the document itself, but in some cases archive maintainers could point the URLs to abstract pages. In these cases the paper will be discarded.

Once the document is saved in our hard disk, we start the conversion process. First, we check if the full text file is compressed. If that is the case, a decompression algorithm is used. Second, we check the file format. Only PDF and PostScript documents are accepted at the moment. Fortunately both are quite popular formats in Economics. More than 95% of the RePEc documents are in either PostScript or PDF.

The last step is to convert the document from PDF or PostScript to ASCII. For this purpose, we use the software pstotext developed by Andrew Birrell and Paul McJones as part of the Virtual Paper project³.

3.2.- Parsing

Parsing is the most complicated process. Authors usually construct references in a variety of formats, even within the same paper. In addition disciplines vary with respect to the traditions in the way citations are marked in the documents.

Due to the importance of the parsing process we decided to start with a software that has been already tested rather than develop new software from scratch. Our choice has been CiteSeer by S. Lawrence, Kurt Bollacker and C. Lee Giles.

CiteSeer is able to identify the part of the document containing the list of references. Then it can split the list into different references. Finally it parses each reference to find the elements. At the moment it only identifies the publication year, the title and the authors. However, as we will see, these four elements are enough for our purposes.

3.3.- Linking

Once we have parsed the documents, the next stage is to look if some of the references successfully found go to documents identified in RePEc. In such cases, some type of link between both documents should be established. We are doing that by comparing each reference successfully parsed, with the authoritative metadata stored in the CitEc knowledge base. At the moment we consider that a reference represents a RePEc document when:

- The parsed reference title and the title in our metadata collection are close enough.
- The publication year of both items is the same.

In this process we take each reference, extract the parsed title and convert it to a normalised version called key title. Here all multiple spaces and articles are removed and upper case letters are converted to lower case. Then we select from our knowledge base of metadata all documents that contain in their title all the words of the reference key title. All selected papers are suspect of being the cited document. In a second step we compute the Levenshtein distance of each suspect title with the reference title. If this distance is greater than 8% of the title length, the suspect document is rejected. Finally, we check if the publication year of the suspect papers and the reference is the same. If this is the case we assume that the reference is to the document we have.

4.- Internal Evaluation

In this section we provide a detailed description of errors detected in the processing of RePEc documents in order to determine if our autonomous citation index could be used to provide bibliometric indicators or to assist in information retrieval.

In order to evaluate the system behaviour we define a series of stages in the reference extraction and linking process that every paper should pass. In this way, the initial stage for all documents is "notprocessed". It will be changed to the final stage of "linked" for papers which have successfully passed all stages in the reference linking process. If the process fails, an error status describing the problem detected is associated with the document. All information about document status and errors is recorded in the knowledge base.

The current version of the system is dated August 2004. It contains 175452 metadata records with information about electronic documents. Such documents are distributed in the 1591 series coming from 378 institutions worldwide contribute to RePEc. The number of documents per institution ranges from those that only provide one or two documents to those with a national scope which provide documents coming from several institutions.

We face three problems when downloading documents. Firstly, there are institutions that charge for access to the full text of their publications. In such cases the documents are simply ruled out. We found 51418 documents with restricted access. That cut down the number of documents to be processed to 124034. The lion share of the restricted documents comes from commercial publishers and the JSTOR project. Secondly, we found in the metadata wrong URLs to the documents' full text. That means the documents are not found at the specified location due to an error in the metadata provided by the archives. Finally, with the error "baddocument" there is a variety of problems. For instance, documents digitalized as images, even using the PDF format, and URLs that instead of pointing to the document's full

text go to an abstract page. This practice is not allowed in RePEc but some institutions work in this way to make sure their web sites get as much hits as possible. Table 1 shows the ratio of not available documents.

TABLE ONE – Documents' availability

Error	Documents	
Restricted	51418	28%
Not found	4221	2%
Bad document	7702	4%
Available Documents	121111	66%

Once the 121111 available documents have been downloaded the system starts the conversion from the original PDF or PostScript formats to ASCII. In this processing step we found three possible problems: "incompatible format" when the format of the file containing the paper is not PDF or PS, "conversion error" when the program that convert the formats fails or "no references" when even having a text version of the paper, the system has been unable to find a references section in the corpus of the document.

61474 documents out of 121111 have been successfully converted to text format. Table 2 shows the error distribution for this particular stage.

TABLE TWO - Errors found in the conversion process

Error	Documents	
Conversion error	10304	9%
No English	2062	2%
No references	24663	22%
Incompatible format	13708	12%
Converted Documents	61474	55%

At this step it is important to note the large number of documents in which the process of conversion has failed. An initial conclusion to be taken into account in future system updates would be the need of testing new conversion programs.

61474 documents were parsed in order to locate and identify their bibliographic references. Documents in which the number of references identified by the system is greater than seventy are discarded. In such cases is quite probably the process has failed since such a large bibliography is unusual. As it is shown in table 3 almost the 90% of documents stay below the limits.

In total, 1165075 references have been identified in 53201 documents. That represents an average of 22 references by document.

The linking module is the last one in the process. It is in charge of creating a link between each reference correctly parsed and the full text of the document it represents if such document is available in RePEc. 307094 out of the 1165075 references identified are representation of RePEc documents.

In conclusion, 44% of documents available in RePEc were successfully processed. That is, the system was able to extract and link their references. More than half of the documents could not be linked for different reasons. The most important cause of problems is the conversion from PDF to text formats. The second most important is that the system was unable to find the references list in the 12% of the documents. Since it is unusual to find an scientific document without bibliography, we could conclude that the algorithm of analysis needs to be considerable improved in order to extract bibliometric indicator with enough quality to be used in bibliometric studies.

TABLE THREE

Error Frecuency		
wrong number of references	2081	13%
correctly parsed	53224	87%

5.- Conclusion

To sum up, along this paper we have described a system that makes possible to automatically extract citation data of documents from a distributed digital library. We have designed a procedure to automatically retrieve the documents' full text from the servers and extract the citation data. Whereas this procedure has been proved successful, a few remarks should be taken into account for future work:

- The collaboration of archives maintainers is a key point to allow a correct administration of the system. Good metadata is essential to obtain relevant results in the citation linking process. The main problem we face is wrong URLs to the documents' full text. This will make impossible to analyse the documents. To solve that we are planning to automatically inform the maintainers about each document that could not be downloaded from their archives.
- Better conversion programs from PDF to text are needed. We work with files generated by a wide range of applications. It is possible even to find scanned documents saved as simple images. As a result we need to use a powerful tool that would allow us to obtain an usable text representation of the document.
- The parsing algorithm could be clearly improved. In our example it was able to parse correctly only the 75% of the references. While this can be an acceptable rate of errors when working with reference linking, it is not enough to create more complicated applications like the bibliometric analysis of a discipline.

6.- References

Krichel, T. Guildford protocol. <ftp://netec.mcc.ac.uk/pub/NetEc/RePEc/all/root/docu/guilp.html>.

Krichel, T. Redif (research documents information format).
ftp://netec.mcc.ac.uk/pub/NetEc/RePEc/all/root/docu/redif_1.html.

Lawrence, S., K. Bollacker, and C. L. Giles (1999). Indexing and retrieval of scientific literature. In Eighth International Conference on Information and Knowledge Management, CIKM99, pp.139_146.

-
1. The precise detail of this base is beyond the scope of this paper.
 2. Linking of documents out of our data set, using technologies based in DOI identifiers, in the same way it is done in CrossRef at the moment, is in our to-do list.
 3. <http://www.research.compaq.com/SRC/virtualpaper/home.html>

Author Information

Mr. José Manuel Barrueco Cruz
Social Science Library, University of Valencia
Campus dels Tarongers, s/n, 46071
Valencia
Spain

barrueco@uv.es
Jose.Barrueco@uv.es
<http://openlib.org/~barrueco>

Trend evaluation and comparison of the use and value of GL in core demography and computer science journals

Rosa Di Cesare, Roberta Ruggieri

CNR – Istituto di Ricerche sulla Popolazione e le Politiche Sociali

Silvia Giannini, Stefania Biagioni

CNR – Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

Abstract

Over the last ten years the impact of grey literature on conventional literature has frequently been studied. Studies have made use of bibliometric instruments used for citation analysis. Recently, this research has magnified attention on the impact of new forms of GL that have emerged along with the spread of Internet.

This work aims to a) measure the impact of GL on two different scientific fields; b) describe the characteristics of GL documents cited; c) ascertain any changes in LG impact due to use of the www.

Two years (1995 and 2003) were chosen as illustrative of the situation before and after the growth in the use of the www.

With these aims, bibliographic references have been analysed in publications in two scientific fields for which it is logical to hypothesise a different impact. The publications are three journals of computer sciences included in the Journal Citation Report (JCR) Science Ed., and three journals of demography included in Journal Citation Report (JCR) – Social Science Ed..

The three journals in each of the two categories were chosen on the basis of their stability during the observation period (1995 and 2003) in terms both of their Impact Factor (IF) - high, medium and low - and of their ranks.

1. Introduction

Over the last ten years the impact of GL on conventional literature has frequently been studied. (e.g., see Alberani, 1990; Di Cesare, 1994; Pelzer and Wiese, 2003; Cordes 2004). Studies have made use of bibliometric instruments used for the citation analysis. These studies have analysed the GL impact on conventional literature in different disciplinary fields and shown its value as an information source, they have also drawn attention to the usefulness of citation analysis, contributing thereby to highlighting the problems connected with their standardization and with GL producers.

Our study examines the impact of GL on demographics and on a specific field of computer science.

The analysis aims to verify the different impact of GL in the two above disciplinary fields: demography more directed towards "national issues" (Katz, 1999), while the second is not influenced by these constraints. As a consequence they use different information sources and cite different types of GL documents.

The comparison is between two years and aims to a) measure the impact of GL on the above scientific fields; and b) describe the characteristics of GL documents cited; c) ascertain any changes in LG impact due to use of the www.

2. Materials and methods

2.1. Sampled journals

We analysed a sample of journals indexed by the Science Citation Index (SCI) and by Social Science Citation Index (SCCI) and included in the ISI-Journal Citation Report (JCR) under the sub-category "Information systems" of the category "Computer science" (CS-IS, for short), and under the category "Demography" (DEM, for short). All selected journals are also included in two more specialized Indexes. Table 1. All selected journals were suggested by domain experts.

In each of the two categories, the three journals were chosen because their IF (high, medium and low) and ranks were stable during the observation period (1995 and 2003).

Table 1. Sampled journals

DEM Journals	IF 1995 (*)	Rank 1995	IF 2002 (*)	Rank 2002
Demography	1.633	1/17	2.060	2/17
Population Research & Policy Review	0.368	11/17	0.510	10/17
International Migration	0.137	16/17	0.357	14/17

All indexed by ISI, Popindex, CICRED

(*) ISI-JCR Social Science ed., 1995 e 2002; Category: Demography

CS-IS Journals	IF 1995 (*)	Rank 1995	IF 2002 (*)	Rank 2002
ACM Transactions on Information Systems (ACM TOIS)	1.176	4/43	1.385	12/77
ACM Transactions on Database Systems (ACM TODS)	0.829	9/43	0.875	27/77
Information Sciences	0.245	30/43	0.361	54/77

All indexed by ISI, INSPEC, EI Compendex

(*) ISI-JCR Science ed., 1995 e 2002; Category: Computer Science-"Information Systems"

Table 1 compares the Impact Factor (IF) and rank in the "Demography" category and in the sub-category "Information systems" in the selected years.

Data was collected directly from primary sources, that is, the bibliographic references of the articles in the selected journals. For each journal we counted:

- a) the number of "articles" including "notes and commentary", "data and perspectives", "emerging research", and "book review";
- b) the number of references (including GL references) in each article;
- c) the number of electronic GL (e-GL, for short) references in each article containing GL references.

2.2. Definitions and indicators

We define GL according to the York recommendations (York, 1978) and considering, in addition, the following integration to its definition (GL1997; GL1999; GL2003). Besides, we refer to the studies on GL that stressed nature, contents and purpose of GL documents rather than specific modality of production and diffusion (Alberani, 1982; Posnett, 1980; Stock, 2003).

We also consider recommendations contained in "instructions to authors" of sampled journals.

For unclear and incomplete citations we consulted:

- specialized indexes
- Library of Congress and British Library catalogues
- the search engine Google

The Manual of Library Association (MLA styling) was consulted to classify EGL documents.

For each GL citation we examined:

- Type of material (conference proceedings/meeting paper, monographic book, report, preprint, private communication, software/technical documentation, thesis, unpublished); including each e-GL reference to URL or to the electronic version of a document or to a home page;
- Date: year of publication
- Language
- Self-citing: depending on whether one of the authors of the GL document cited was also an author of the article examined.

As to the type of material cited it should be noted that:

- the type "proceedings/meeting papers" includes: conference proceedings that are not published by a commercial publisher; seminars, workshops, posters and communications;
- the type "report" includes: research and technical memoranda, white papers, working papers, illegal reports, project reports, discussion papers, occasional papers, research reports; all of which were numbered by their producers;
- the type "software/technical documentation" includes: software, user guides, manuals, standard, patents and technical documentation of computer programs and for statistical surveys;
- the type "thesis" includes Ph.D. dissertations and theses;
- the type "preprint" includes: preprints, "submitted-to" manuscript, "to-be-published" manuscript, "in press", "forthcoming";
- the type "monographic book" includes all documents with a title, but without reference number;
- the type "unpublished" includes all documents cited as unpublished material and manuscript;
- the type "statistical report" includes official and unofficial statistical reports (census, survey, statistical databases);
- the type "Home page" includes all documents cited using the home page of their producer;
- Information Database (database, e-print archive, digital library)
- the type "miscellaneous" includes all GL materials not belonging to any previous type;
- the type "undefined" includes all documents that were incompletely cited.

We measured the different impact of GL on the two disciplinary fields, using the following indicators:

- the frequency of GL use (i.e. the proportion of articles with GL citation, out of all articles examined); (table 2)
- the frequency of GL citing (i.e. the proportion of GL references out of all the references examined); (table 3)
- the intensity of GL use (i.e. the frequency of GL citing divided by the frequency of GL use) – this is an indicator of the average frequency of GL references per article with GL; (table 4);
- the frequency of e-GL use for each article containing GL references (table 5 and 6)

3. Findings

3.1. General comments

The DEM journals contained 85 articles for a total of 2,822 references in 1995 and 114 articles in 2003 for a total of 4,181 references. In "Notes and commentary", "data and perspectives", "emerging research", and "book review", the number of GL references were not remarkable.

The CS-Information Systems journals contained 112 articles for a total of 2,561 references in 1995 and 167 articles in 2003 for a total of 3,872 references. In CS-IS we found only articles with GL references.

Tables 2, 3 and 4 show:

- for DEM: the frequency of GL use increased, the frequency of GL citing decreased and the intensity decreased;
- for CS-IS the frequency of use remained stable, the frequency of GL citing increased and the intensity of GL increased.

The journal "Demography" with highest IF has the stable value for all the three indicators. For the journal ACM TOIS with the highest IF, only the value of frequency of GL use is stable.

Table 2. Frequency of GL use (1995 vs. 2003)

DEM Journals	1995			2003		
	Number of articles	Number of article with GL	Frequency of GL use (%)	Number of articles	Number of article with GL	Frequency of GL use (%)
Demography	37	37	100.0	38	38	100.0
Popul Res Policy Rev	27	23	85.2	37	36	97.3
International Migration	21	21	100.0	39	38	97.4
Total	85	81	95.3	114	112	98.2
CS-IS Journals						
ACM TOIS	16	15	93.8	15	14	93.3
ACM TODS	12	12	100.0	15	13	86.7
Information Sciences	84	58	69.0	137	100	73.0
Total	112	85	75.9	167	127	76.0

Table 3. Frequency of GL citing (1995 vs. 2003)

DEM Journals	1995			2003		
	Number of references	Number of GL references	Frequency of GL citing (%)	Number of references	Number of GL references	Frequency of GL citing (%)
Demography	1,27	237	18.7	1,795	326	18.02
Popul Res Policy Rev	779	295	37.9	1,088	210	19.03
International Migration	773	413	53.4	1,298	444	34.2
Total	2,822	945	33.5	4,181	980	23.4
CS-IS Journals						
ACM TOIS	618	71	11.5	602	127	21.1
ACM TODS	441	45	10.2	621	58	9.3
Information Sciences	1,502	172	11.5	2,649	434	16.4
Total	2,561	288	11.2	3,872	619	16.0

Table 4 Intensity of GL use (1995 vs. 2003)

DEM Journals	1995		2003	
	IF	Intensity of GL use (%)	IF	Intensity of GL use (%)
Demography	1.633	18.7	2.060	18.2
Popul Res Policy Rev	0.368	44.2	0.510	19.8
International Migration	0.137	53.4	0.357	35.4
Total		35.2		23.8
CS-IS Journals				
ACM TOIS	1.176	11.9	1.385	22.6
ACM TODS	0.829	10.2	0.875	10.7
Information Sciences	0.245	16.7	0.361	22.5
Total		14.28		21.0

e-GL use

Tables 5a and 5b show e-GL use in our sample. As expected we found that the majority of the e-GL citations was concentrated in the CS-IS field.

The availability of information on the web makes the access to documents easier, in particular to software, technical documentation, home page, digital libraries, e-print archive, data base etc., therefore, in CS-IS authors started linking citations to on-line documentation earlier than the authors in DEM. However, the delayed use of e-GL in DEM is counterbalanced by its higher frequency in 2003: 35,7% vs. 32,3% (Table 5b).

As shown by the last columns in Table 5b) a large proportion of e-GL documents for both DEM and CS-IS are still accessible in 2004 (70.6% and 82.4%).

Table 5a. Frequency of e-GL use (1995)

	1995			Availability	2004
	Number of GL references	Number of e-GL references	Frequency of e-GL use (%)	Number of e-GL available	(%)
DEM Journals					
Demography	237	0	0.0	0	0.0
Popul Res Policy Rev	295	0	0.0	0	0.0
International Migration	413	0	0.0	0	0.0
Total	945	0	0.0	0	0.0
CS-IS Journals					
ACM TOIS	69	9	13.3	2	22.2
ACM TODS	45	2	16.7	1	50.0
Information Sciences	172	0	0.0	0	0.0
Total	286	11	4.7	3	27.3

Table 5b. Frequency of e-GL use (2003)

	2003			Availability	2004
	Number of GL references	Number of e-GL references	Frequency of e-GL use (%)	Number of e-GL available	(%)
DEM Journals					
Demography	326	32	42.1	25	78.1
Popul Res Policy Rev	210	12	16.7	5	41.7
International Migration	444	58	47.4	42	72.4
Total	980	102	35.7	72	70.6
CS-IS Journals					
ACM TOIS	127	49	78.6	40	81.6
ACM TODS	58	17	38.5	16	94.1
Information Sciences	434	87	25.0	70	80.5
Total	619	153	32.3	126	82.4

3.2. Characteristics of the GL documents.

3.2.1. Form of materials cited in both fields.

In the previous sub-section we found that the GL use decreased in DEM and increased in CS-IS. Table 7 reports the distribution of GL documents by document type and shows that in DEM the only document that increased are: technical document, home page, db, dl, archive and undefined and in CS-IS the only document type that increased are: proceedings/meeting papers, technical document, miscellaneous, home page, monographic book, db, dl, archive.

It should be noted that from 2003 both disciplines have specific e-GL references to Home Pages.

Table 6a. DEM distribution of GL documents by document type (1995 vs. 2003)

DEM	1995		2003	
Document type	Number	(%)	Number	(%)
Thesis	16	1.7	30	3.1
Report	189	20.0	183	18.7
Preprint	0	0.0	0	0.0
Proc/Meeting p.	169	17.9	116	11.8
Unpublished	26	2.8	13	1.3
Technical Doc	19	2.0	21	2.1
Miscellaneous	26	2.8	53	5.4
Undefined	3	0.3	27	2.8
Home Page	0	0.0	25	2.6
Monographic book	233	24.7	236	24.1
Statistical report	264	27.9	267	27.2
Db, DI, Archive	0	0.0	9	0.9
Total	945	100.0	980	100.0

Table 6b. CS-IS distribution of GL documents by document type (1995 vs. 2003)

CS-IS	1995		2003	
Document type	Number	(%)	Number	(%)
Thesis	43	14.9	69	11.1
Report	85	29.5	119	19.2
Preprint	25	8.7	19	3.1
Proc/Meeting p.	72	25.0	238	38.4
Unpublished	5	1.7	1	0.2
Technical Doc	47	16.3	110	17.8
Miscellaneous	3	1.0	12	1.9
Undefined	7	2.4	11	1.8
Home Page	1	0.3	23	3.7
Monographic book	0	0.0	8	1.3
Statistical report	0	0.0	0	0.0
Db, DI, Archive	0	0.0	9	1.5
Total	288	100.0	619	100.0

3.2.2. Distribution of GL documents by date.

Table 7a reports the distribution of GL documents in DEM by date. We found that more than 50% of the citations were concentrated in the last five years for 1995 and in the last four years for 2003, with maximum value in 1993 and in 2001.

Table 7a. Distribution of GL documents by date in DEM

1995			2003		
Date	Frequency	Cumulative frequency (%)	Date	Frequency	Cumulative frequency (%)
1995	9	1.0	2003	40	4.1
1994	84	9.8	2002	135	17.2
1993	150	25.7	2001	141	32.2
1992	100	36.3	2000	107	43.2
1991	94	46.2	1999	96	53.0
1990	97	56.5	1998	57	58.8
1989	51	61.9	1997	67	65.6
1988	63	68.6	1996	44	70.1
1987	29	71.6	1995	44	74.6
1986	29	74.7	1994	30	77.6
1985-1976	166	92.3	1993-1984	151	93.1
1975-1966	31	95.5	1983-1974	32	96.3
1965-1956	10	96.6	1973-1964	17	98.1
1955-1900	21	98.8	1963-1900	6	0.1
[s.d]	11	1.2	[s.d]	13	1.3
Total	945		980	100.0	

Table 7b. Distribution of GL documents by date in CS-IS

1995			2003		
Date	Frequency	Cumulative frequency (%)	Date	Frequency	Cumulative frequency (%)
1995	18	6.3	2003	17	2.7
1994	24	14.6	2002	43	9.7
1993	35	26.7	2001	84	23.3
1992	41	41.0	2000	57	32.5
1991	49	58.0	1999	59	42.0
1990	23	66.0	1998	51	50.2
1989	14	70.8	1997	50	58.3
1988	12	75.0	1996	35	64.0
1987	10	78.5	1995	40	70.4
1986	5	80.2	1994	21	73.8
1985-1976	29	90.3	1993-1984	78	86.4
1975-1966	8	93.1	1983-1974	13	88.5
1965-1956	1	93.4	1973-1964	6	89.5
1955-1900	0	0.0	1963-1900	0	0.0
[s.d]	19	6.6	[s.d]	65	10.5
Total	288		619		

In CS-IS we found more than 50% of the citations were concentrated in the last five years for 1995 and 2003 with maximum value in 1991 and in 2001. (see table 7b).

3.2.3 Self-citing and language

Self-citing: in both disciplinary fields the proportion of self citations decreased. Note that the journals with low IF seem to be more affected by the phenomenon of self-citing.

Language. In CS-IS GL documents cited were written all in English and in DEM they represent more than 90%.

Table 8. Distribution of self-citations (1995 vs. 2003)

DEM Journals	1995			2003		
	Number of GL references	Number of Self-cit.	(%)	Number of GL references	Number of Self-cit.	(%)
Demography	237	21	8.9	326	14	4.3
Popul Res Policy Rev	295	13	4.4	210	12	5.7
International Migration	413	52	12.6	444	25	5.6
Total	945	86	25.9	980	51	15.6
CS-IS Journals						
ACM TOIS	71	17	23.9	127	22	17.3
ACM TODS	45	12	26.7	58	12	20.7
Information Sciences	172	68	39.5	434	99	22.8
Total	288	97	33.7	619	133	21.5

CONCLUSION

We applied a methodology to two scientific fields, which are far from each other for citation models in order to discover the impacts of GL and e-GL in two different years, one before and the other after the introduction of electronic communication technologies.

We found that in DEM the use of GL tends to decrease and in CS-IS tends to increase also thanks to e-GL. This variation in the impact of GL is independent from the characteristics of GL document types, both in the case of increase (CS-IS) or decrease (DEM) of its impact.

Since the quality of GL citations were unclear and incomplete, the analysis was time-consuming. Moreover we were quite surprised to observe that the majority of journals analysed does not contain recommendations in the "instruction to authors" which facilitate the correct citation of GL and e-GL. Nevertheless we think that examining the bibliographic references, we may have further insight on the impact of GL and e-GL in different knowledge fields, thereby achieving at the same time deeper understanding of the increasing overlapping of both electronic GL and conventional literature.

REFERENCES

- Alberani W., De Castro P. (2000). The use of free literature in health sciences: a preliminary survey results. *Bullettin of Medical Library Association*, 78 (4), 358-363.
- Cordes Ruth (2004). Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an International Marine Scientific Advisory Body. *The Canadian Journal of Information and Library Science*, 28 (1) 2004.
- De Blaaij Cees. Grey literature from invisibility to visibility: the quest for grey content in the domain of the invisible web., in GL5, Fifth International Conference on Grey Literature, Amsterdam, December 2003.
- Gelfand Julia. 2003. The ubiquity of grey literature in a connected content context. in GL5, Fifth International Conference on Grey Literature, Amsterdam, December 2003
- GL97, Third International Conference on Grey Literature, Luxemburg, GreyNet, November 1997.
- GL99, Fourth International Conference on Grey Literature, Washington, GreyNet, December, 1999.
- GL5, Fifth International Conference on Grey Literature, Amsterdam, TextRelease, December, 2003.
- Stock C., Schopf J., Grey literature in an open context: from certainty to new challenges, in GL5, Fifth International Conference on Grey Literature, Amsterdam, December 2003
- Luzi Daniela, Internet as a new distribution channel of scientific grey literature. The case of Italian WWW servers, in GL2, Second International Conference on Grey Literature, Washington, December, 1995
- Katz J. Sylvan.. Bibliometric indicators and the Social Sciences, prepared for ESRC. SPRU - University of Sussex . June 2004 <http://www.sussex.ac.uk/Users/sylvank/pubs/ESRC.pdf>
- Mackenzie Owen, John. The expanding horizon of grey literature. October 2004.
<http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/glpaper.htm>
- MLA. Using MLA style to cite and document sources. June 2004.
<http://www.bedfordsmartins.com/online/cite5.html>
- Nancy L. Pelzer, M.A., Wiese, William H. (2003). Bibliometric study of grey literature in core veterinary journals. *Journal of Medical Library Association* 91 (4).
- NIDI, The world wide web of demography. June 2004.
www.nidi.nl/links/nidi6000.html

Author Information

Ms. Rosa Di Cesare
Istituto di Ricerche sulla Popolazione e le Politiche Sociali, IRPPS/CNR
Via Nizza, 128
Rome 00198
Italy

biblio.irp@irpps.cnr.it
www.irpps.cnr.it

Impact of the Inclusion of Grey Literature on the Scholarly Communication Patterns of an Interdisciplinary Specialty

Kathel Dunn, MSLS
Ehrman Medical Library, New York University School of Medicine

Abstract

Problem/goal: Researchers who more routinely use traditional scholarly literature published in peer-reviewed journals are especially disadvantaged in understanding, accessing and obtaining grey literature in their fields of research. (Lasker, 1998) For example, researchers in health policy find that their research depends on both the well-organized and relatively easily accessible journal literature used by scientists and clinicians; the researchers also need to find reports literature published by think tanks and independent research organizations. The purpose of this research is to identify the interconnectedness – or not – of health care policy research through a citation analysis of journal literature and grey literature.

Research methodology: This research is a citation analysis study of health care policy, expanding the literature studied to include not only the journal literature used by researchers but also the grey literature produced by think tanks and independent research organizations. The working hypothesis of this paper is that the evidence will bear out anecdotal comments that there is a division or separateness between the two types of literatures; and also, that the literatures are seemingly inaccessible to the researchers and users of the different groups. That is, that the literature most used and identified with Group A is difficult to locate and less likely to be used by Group B and vice versa.

The study will analyze the references/citations of journal articles to both journal articles and grey literature and the references/citations of grey literature to journal articles and grey literature published in the last five years on changes in Medicaid, a health care policy issue. The purpose of the study is to assess the impact that the inclusion of grey literature has in a citation analysis of a particular field.

Inclusion of grey literature may have no impact on the citation-generated network of a particular field, serving only to strengthen the already existing network or image. Or, inclusion of grey literature may have a strong impact on a network, revealing parts previously invisible when using only the journal literature. The revealed, or more complete network picture, may indicate that both literatures are well-integrated, each citing the other to a high degree. Or, the complete network picture of a field may indicate that the literatures instead exist in parallel, each communicating on the same topic, but seemingly – as indicated by the citation analysis – not communicating to, or for, each other. If the literatures exist side by side, and are not well-integrated – as indicated by citations – this raises a series of questions for health policy researchers. Is the lack of shared citations a sign that one of the literatures does not cite? Or does the lack of shared citations mean that the literatures do not have significance for each other?

Anticipated results: It is anticipated that the journal literature will have more references, more references to journal literature and fewer references to grey literature (specifically reports literature). Grey – reports – literature will have fewer references than journal literature and when the grey literature does cite it will more frequently cite non-journal literature.

Introduction

Researchers who more routinely use traditional scholarly literature published in peer-reviewed journals are especially disadvantaged in understanding, accessing and obtaining grey literature in their fields of research (Lasker, 1998). For example, researchers in health policy find that their research depends on both the well-organized and relatively easily accessible journal literature used by scientists and clinicians; the researchers also need to find reports literature published by think tanks and independent research organizations. The purpose of this research is to identify the interconnectedness – or not – of health care policy research through a citation analysis of journal literature and grey literature.

This research is a citation analysis study of health care policy, expanding the literature studied to include not only the journal literature used by researchers but also the grey literature produced by think tanks and independent research organizations. The working hypothesis of this paper is that the evidence will bear out anecdotal comments that there is a division or separateness between the two types of literatures; and also, that the literatures are seemingly inaccessible to the researchers and users of the different groups. That is, that the literature most used and identified with Group A is difficult to locate and less likely to be used by Group B and vice versa.

Research in the area of specialties and in the related areas of invisible colleges, research fronts, epistemic communities, etc has focused on first identifying and documenting the existence of a specialty. A subsequent line of research worked at characterizing the nature of the specialty and its interactions:

whether its members are composed of the geographically close or distant; whether its members are socially connected or institutionally connected and the extent to which its members related to each other and each others' work. In most instances, analysis of specialties' written record of research used the traditional scholarly mediums of books and journals (journal articles). Books and journals were used in citation analysis, one of the most common methodologies in scholarly communication research available to outline, define and discover specialties. Journal literature was a natural choice for study as journals were a clear well-organized communication medium of scholars and researchers. Journal literature is highly visible, and easily counted whether through library-use studies or through a bibliographic or citation database.

The explication of specialties has often used specialties that focus on one area, that bring a natural coherence to their subject domain. Examining an interdisciplinary field brings another set of challenges with it. Where one field may communicate to each other by way of meetings, conference abstracts and preprints, another may depend on the journal article. When each of these fields meet – or don't – around an issue, communication represents a particular challenge. Individual scholars may make deliberate choices to publish outside their field to make their ideas and scholarship known to other researchers, but that method presupposes in part that the choice of communication tool is a valid one within that other field. The specialty of health care policy is an interdisciplinary field of, at the least, researchers in medicine, administrators and policy makers at the local, state and federal levels of government. Where researchers in medicine most often communicate to each other through the journal article, researchers in policy depend on a broader set of communication medium including think tank publications (grey literature), government reports (grey literature) and the media.

Where grey literature in academe may have a history of being considered "draft-like", awaiting its day as a peer-reviewed publication, for policy researchers grey literature is usable as is; in fact its final form as a report requires no other public peer-review. When researchers who publish in journals write on the same topic as researchers who publish in think tank or independent research organizations, how do these two sets of literatures interact, if at all?

Grey literature on the other hand has less frequently been included in analyzing specialties, and in particular in citation analyses that define and chart a specialty. There are a number of reasons for this: grey literature is elusive, it lacks a clear definition (and thus a clearly operationalized variable), and it is not the formal medium of communication of the academic establishment, and thus perhaps not as visible a medium for study. The lack of inclusion of grey literature in studies was perhaps influenced by the focus of scholarly communication on the peer-reviewed journal literature and by extension the users' use of it.

Literature Review

In both public and private statements, researchers and policy makers indicate that knowing or finding the literature they need for their work is difficult. This suggests a number of possibilities. It could be that the available literature – in whatever form – is not well-represented in channels where they can be found: in bibliographic databases or abstracting services or by other means made clearly accessible to its audience. It could be that the literature desired by one group (academic researchers) is written by another group (think tank researchers) for another group entirely (policymakers). Or, a literature written by academic researchers is not easily accessible to another group (policymakers or think tank researchers) who may wish to make use of it. Similar research in a given research field may take different forms – journal article, think tank report, etc – depending upon the intended audience.

There is at least some anecdotal evidence to suggest that specialties exist but that researchers' work is published in isolation or if the researchers are known to each other socially (informally), the formal tools available may not make the work visible.

The conveners of one conference on the challenges of accessing information in public health and health policy cite the need for grey literature, particularly in the fields of health policy and public health: "This conference, *Accessing Using Information: Challenges in Health Policy and Public Health*, grew from the sense of the organizers that the bibliographic resources available to researchers and practitioners in the fields of public health and health policy may not meet the needs of the users as well as they might and do not match the resources available to biomedical researchers and practitioners" (Gray, 1998). In one small study presented at the conference, researchers examined the literature needed for public health / health policy researchers. They found that while the coverage for biomedical literature is well-known and available in databases, the (grey) literature necessary for public health and health policy researchers is not as well-covered by health and policy databases. Other researchers have held their own conferences that focus on their frustration in locating literature (most often grey) that they see as vital to their own work (Cookson, Archard, & McDaid, 1999; Gray, 1998; Lasker, 1998).

From the policy makers' side of the question, it appears that the journal literature is not satisfying. In an October 2003 interview in the New York Times magazine on his plans to create a think

tank to generate liberal, Democratic policy ideas, John Podesta, former Chief of Staff in the Clinton White House, discussed the need for a "project to address long-term challenges facing public education, relying on academics who may have innovative solutions. 'I'm sure there are journals filled with that stuff, that test and analyze and explain it," Podesta said excitedly, "but no one comes forward to put that into the policy arena'" (Bai, 2003).

And in a *Health Affairs* (journal) article, Saul Feldman laments, "Like strangers in the night, dimly aware of each other's presence, health services researchers and managed mental health organizations have been exchanging glances but not much more. As a result, research findings have had only a negligible effect on managed mental health care". He continues, "And so the funded studies, meticulously done by well-respected researchers and methodologically correct, qualify for publication in journals that, by and large, are read by other researchers, not by those with power over who gets what kind of mental health services" (Feldman, 1999).

The purpose of this proposal is to examine the specialty of health care policy in Medicaid and expand the literature studied to include not only the journal literature used by university-based academicians but also the grey literature produced by think tanks and independent research organizations. The working hypothesis of this paper is that the evidence will bear out the anecdotal comments that there is a division or separateness between the two types of literatures; and also, that the literatures are seemingly inaccessible to the researchers and users of the different groups. That is, that the literature most used and identified with Group A is difficult to locate and less likely to be used by Group B and vice versa.

A note on grey literature

One type of grey literature is reports literature. Reports literature pre-dates grey literature. Reports are documents published by think tanks or independent research organizations. Unlike preprints, conference proceedings or dissertations and theses, reports are usually not published again in another format. The question with reports literature, given that it is written outside the traditional scholarly communication channels, and shows no sign of movement from its grey form to another non-grey form, is how to understand its role and importance to academe, beyond anecdotal reports and complaints.

In some fields of study, more than university-based academicians are involved in a research field. Government, independent research and think tanks all are involved in a variety of policy issues that impact and influence research in education, health and transportation. Government, independent research and think tanks often publish not in the traditional peer-reviewed journal literature but in independent, self-published (non-commercially published) reports. These reports are grey literature.

Citation analysis and the development of scientific specialties

Early work in citation analysis identified networks, groups, research fronts (Mulkay, Gilbert, & Woolgar, 1975; Mullins, Hargens, Hecht, & Kick, 1977; Price, 1965; Price & Beaver, 1966). The citation analysis was usually conducted in a field with only one type of publication: the journal article. Journal articles had become eminently countable through the use of citation databases from the Institute of Scientific Information, and this then became a fruitful tool in scholarly communication research. The methodology was frequently picked up and used by researchers outside the field of information and library science and used by those in the subject areas themselves, almost like a professional novelty act. What the studies were good at identifying were publications by like individuals with similar purposes. Such studies provided a pictorial depiction – through citation maps – of the activities of scholars. They might, for example, show how a field developed and matured, and point toward future trends. The studies also could identify who was working in a field and distantly connected researchers and work.

Citations in a particular field also charted the growth and development of a particular field (development of scientific specialties): its emergence, full development and its breaking off into other fields or specialties.

Any citation analysis of a field such as health policy that does not include grey literature is by its nature incomplete. Certainly a citation analysis of journal articles in health policy can identify a research group, chart the emergence, growth and further specializations in the field but the picture it draws will inevitably be incomplete. The picture painted by journal citation analysis in health policy that includes only the peer-review journal writing academicians excludes the work of think tanks, independent research and government organizations.

Methodology

A search for journal articles was conducted in the ISI database Web of Science on "Medicaid" and "reform". The database covers three indexes, Science, Social Science and Arts & Humanities Citation. In total the database fully indexes 8,700 journals and selectively indexes an additional 10,100 journals. The search was limited to journal articles published since 2000.

The search for reports literature was conducted by reviewing the web sites of grey literature producing organizations ("think tanks"), known to research and publish in the field of health and health policy. For the purposes of this study, the list of grey literature producing organizations were culled from a list of grey literature "publishers" obtained from a group of health policy/public health researchers who had previously expressed an interest in identifying grey literature in health, health policy and public health. The researchers were given a short list of known grey literature publishers and asked to expand on that list suggesting other institutes, associations and government organizations that produce grey literature and would be used in their own work. The think tank web sites were either searched for the word "medicaid" or their publication lists were browsed. Publications from 2000 on were included.

Results

The search for journal articles returned 73 citations from 47 unique journal titles. The most articles on Medicaid reform were in the journal, *Health Affairs*. The top 13 journals by articles on the topic "Medicaid reform" are listed in Table 1.

Table 1. Top Journals, with articles on Medicaid reform

No.	Journal	Articles	% Of Total
1	Health Affairs	10	13%
2	Journal Of Health Politics Policy And Law	7	9%
3	Gerontologist	4	5%
4	Health Care Financing Review	4	5%
5	Academic Medicine	3	4%
6	Journal Of Economic Perspectives	3	4%
7	Educational Gerontology	2	3%
8	Health Services Research	2	3%
9	Inquiry-The Journal Of Health Care Organization Provision And Financing	2	3%
10	Jama-Journal Of The American Medical Association	2	3%
11	Journal Of General Internal Medicine	2	3%
12	Journal Of Health Economics	2	3%
13	Oncology-New York	2	3%

The search for grey literature – reports – produced 79 reports from 17 think tanks or consultant groups. Because the consultant groups would conduct research and publish for a think tank, and because of the multiplicity of connections between and among the organizations, the organizations are listed in alphabetical order only in Table 2, with no assigned number of reports for each one. For the purposes of this study, the grey literature reports included research reports, survey reports and literature reviews. Where possible, reports that were opinion pieces or published in serial format (and thus could be considered a journal or periodical) were excluded.

The journal articles had a total of 2,690 citations, with an average of 37 citations per article. The reports literature had a total of 1,163 citations, with an average of 15 citations per report. Each citation was examined and then categorized by format: journal article, grey literature report, government document, book or other (non-categorized). Journal articles cited other journal articles more than any other format type; think tank reports cited think tank reports more than any other format type.

Table 2. Citations by format type

	Journal articles	%	Think Tank reports	%
Journal articles	1536	57%	249	21%
Think Tank reports	495	18%	578	50%
Government documents	305	11%	264	23%
Books	257	10%	13	1%
Other	97	4%	59	5%
References (total)	2690		1163	

Discussion

Alberani et al. found in their study that 67% of articles representative journals in the health sciences cite grey literature (1990). In this study, 84% of the journal *articles* cite grey literature. In contrast, 61% of the think tank *reports* cite journal articles. Journal articles about Medicaid reform show a marked awareness of think tank reports (grey literature). While these reports may not be found in traditional abstracting and indexing databases, the researchers and policymakers in Medicaid reform can and do locate them.

It appears that at least through a citation analysis, that there is a cohesive network of authors and researchers in the field of Medicaid reform. Though researchers complain of a lack of suitable tools to access grey literature and other resources outside of the biomedical world, it seems that grey literature is available to them. More striking, however, is the little use of journal articles by reports authors. It appears that the comment that good ideas are “locked up” in journals may be true, especially in an open-access web environment, where many of the grey literature reports are freely available for reading, printing and downloading and the journal articles require subscription access.

There are limits to citation analysis. Author intent, self-citation, time constraints, are all factors that might confound the data. This study does suggest, though, that grey literature (reports literature) has become much more accessible to its users and implies that journal literature, while still heavily cited, is less used by policy advocates and policy makers. The interdisciplinary specialty of Medicaid reform is weighed towards the think tank authors of reports literature.

References

- Alberani, V., Pietrangeli, P. D. C., & Mazza, A. M. R. (1990). The use of grey literature in health sciences: A preliminary survey. *Bulletin of the Medical Library Association*, 78(4), 358-363.
- Bai, M. (2003). Notion building. *New York Times Magazine*, 82-87.
- Cookson, R., Archard, L., & McDaid, D. (1999). Electronic bibliographic tools for incorporating social science research into health care must be improved. *BMJ*, 319(7204), 257-258.
- Feldman, S. (1999). Strangers in the night: Research and managed mental health care. *Health Affairs*, 5(18), 48-51.
- Gray, B. H. (1998). Sources used in health policy research and implications for information retrieval systems. *Journal of Urban Health*, 75(4), 842-852.
- Lasker, R. D. (1998). Challenges to accessing useful information in health policy and public health: an introduction to a national forum held at the New York Academy of Medicine, March 23, 1998. *Journal of Urban Health*, 75(4), 779-784.
- Mulkay, M., Gilbert, G. N., & Woolgar, S. (1975). Problem areas and research networks in science. *Sociology*, 9, 187-203.
- Mullins, N. C., Hargens, L. L., Hecht, P. K., & Kick, E. L. (1977). The group structure of cocitation clusters: a comparative study. *American Sociological Review*, 42(4), 552-562.
- Price, D. S. (1965). Networks of scientific papers. *Science*, 149, 510-515.
- Price, D. S., & Beaver, D. B. (1966). Collaboration in an invisible college. *American Psychologist*, 1011-1018.

Author Information

Ms. Kathel Dunn,
Associate Director for Public Services
Ehrman Medical Library
New York University
550 First Avenue
New York, NY 10016
USA

Email: dunn@library.med.nyu.edu
dunnk01@library.med.nyu.edu

Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit

J. Schöpfel and C. Stock, INIST/CNRS, France

D.J. Farace, GreyNet, Netherlands

J. Frantzen, Boer & Croon, Netherlands

Introduction

The goal of this research project is to learn more about the expanding field of information known as 'grey literature'. Citation analysis allows one the possibility to follow the work of the authors in the GL-Conference series, as well as authors whom they have cited both in grey and commercial publishing. A further goal of this paper is to examine the value of the GL- Conference Proceedings for research in the field of grey literature by examining among other things, its impact on the work of contributing authors. This study was sparked by a general interest in citation analysis as an instrument used in various sciences including information science. And, in particular by a paper presented at GL5 in which grey literature as a main channel for publication by an International Marine Scientific Advisory Body was investigated (MacDonald et al., 2004).

The stakeholders in this study are not first and foremost the authors who produce a document type, which falls within the category of grey literature. Instead this study is directed to those authors who focus their research and writing on the topic of grey literature, these are referred to as the meta-authors on grey literature. They have entered the field of information studies and focussed on a particular area, they entered in small numbers and in the past decade they have become contending stakeholders.

In the first five volumes of the GL Conference Proceedings a total of 139 conference papers were published, 30 of which did not contain references. Bibliographic data shows that 152 authors/co-authors, which is a full sample, were responsible for the content of these papers. 108 are 1st authors, 78 of whom are sole authors. 30 of the authors (>20%) have published more than once in the conference series and 6 of whom (>.05%) have published three or more times in this series. It is important that those authors who published within the series are also published outside the series – indicating that their clout extends beyond the grey circuit. In this way, their sphere of influence covers their entire work and reflects well on the conference series as a whole. The direction until now has been grey to commercially published e.g. a conference paper is later published as a journal article; however, with the advent of OAI, Open Archive Initiative, this direction could now even allow for reversal.

Procedure and Method

This research project is divided into three stages:

1. The first stage is the compilation of data from the first 5 volumes in the GL Conference Proceedings (1994-2004) in which the parameters are defined, software chosen, the fields described, a division of labour among the project workers, creation of a database, control of the input, and an informal documentation of the process by way of a project log.
2. The second stage is a literature search on citation analysis previously carried out in the field of grey literature and to some extent beyond.
3. The third stage is the analysis of the primary data from Stage 1 within the framework of the findings from the literature research in Stage 2, as well as a reflection on the problems encountered and the observations made during the first two stages. This final stage also includes the writing of the paper and the creation of a PowerPoint presentation.

Data Collection and Processing

The printed and published volumes of the GL proceedings served as the primary source for our data collection. Since there was no electronic version of the earlier volumes available, the data input had to be done manually. The fact that there was no standard applied to citations from the onset of this series meant that a manual input would have been required even if the e-papers were available. MS Access was the most used database software among the four project workers; therefore this was the most suitable choice. The definition and description of the fields in the database, which would constitute a record was the next task ahead. The final number of fields was 19 and the final number of records was 1374, which brought the total number of possible data entries to 26106. This is also counting the number of fields left blank. For example up to four fields were available for Cited Sources i.e. 1st through 4th author. When there was only one Cited Source the other three fields were left blank. When the comparative analysis was carried out using the index from the Annotated Bibliography on the topic of Grey Literature (4th ed., 2000) the possible number of entries in the Citation Database increased to 31602.

The table below provides an overview of the record fields and a description of each field. The description was important for the data input since four project workers were entering data simultaneously on multiple locations. This however did not detract from the fact that fields would later have to be expanded and data entry corrected and standardized.

RECORD FORMAT AND FIELD DESCRIPTIONS

Field Name	Description of the Field
Record Number	Record Numbers were assigned after the final record input starting with 1001 ranked in the order of (1 st) Date of Publication, (2 nd) Sourced by, and (3 rd) Citation Number
Series Number	Number of the Proceedings in the GL-Conference Series from 1 to 5
Date of Publication	Date the Proceedings were published not the date of the conference
Sourced by	The author(s) of the paper in the conference proceedings (E.g. Carroll, B.C.; Cotter, G.A.)
Type of Citation	Endnote, Footnote, Non-Numbered (Reference) * Later expanded to include No Citations
Citation Numbering	Yes, No, Non-Applicable * This field was later added to all of the records
Citation Number	If numbered, the number as it appears in the paper, whether endnote or footnote
Format of Citation	Standard (i.e. bibliographic) or Hyperlinked * Later expanded to include Explanative Notes as well as Names & Addresses of Organisations cited
Citation Date	Date of the cited work
Serial Citation	Does the citation refer to another paper in the GL-Series * Later expanded to include other GreyNet publications (Yes/No)
Type of Publication	Examples: Book/Monograph, Proceedings, Report, Web-paper, etc. * Later changes were made to achieve uniformity (e.g. article to journal article, website to Webpage, etc.)
Kind of Publication	Commercial Publication, Grey Literature, Uncertain * Later, as a result of the first team meeting, the number of 'Uncertain' entries was significantly reduced
Language	Language of cited work (English, French, Italian, etc.)
Cited Source	Person or Non-Person
Self Citation	Does it refer to one of the Authors/Co-authors of the conference paper: Yes, No, Non-Applicable (i.e. when a non-person is cited) * Later expanded to include publications by ones own organization.
Name of 1 st Cited Source	If a person (e.g. Gelfand, J.) Otherwise use the title of the journal/work * Later expanded to include the Acronym of a Corporate Author (E.g. CEC, CERN, etc.)
Name of 2 nd Cited Source	Ibid, if applicable
Name of 3 rd Cited Source	Ibid, if applicable
Name of 4 th Cited Source	Ibid, if applicable

No doubt, the title of this conference 'Work on grey in Progress' best describes the work that was done on this paper. It allowed the four project workers leeway to implement and legitimate later additions, changes, and expansion of the record fields. In fact, 11 of the 19 fields resulted in this way. In June of 2004, when the team first met for a full day in Amsterdam, the final revisions were adapted. And in October of 2004, when the team met for a second time in Nancy, only textual changes were made in the Description of the fields. The 8 fields that remained unchanged from the outset of the research project are quite straightforward, and require no further explanation at this point. However, the fields that were altered or expanded are dealt with here explicitly. It might be assumed that these 11 fields identify more than the other 8 what unique is, or what sets a citation analysis of grey literature apart from other conventional citation analyses?

Those record fields that were added, revised, and/or expanded during the course of the data collection include:

- **Type of Citation:** Authors, who did not furnish notes or references in their papers, would still be entered in the citation database. And, in the field Type of Citation, the term "No Citations" was entered. In the field Citation Numbering, the term "Non-applicable" was used. The other fields in these records were then left blank. While this accounts for only 30 of the 1374 records (0.02%) in the database, it amounted to 30 of the 139 papers (21%) in the GL-Series
- **Citation Numbering:** The field Citation Numbering was later added bringing the total number of fields to 19 per record.
- **Format of Citation:** To the field Format of Citation were added the terms: "Explanative" to indicate an author's note that was given outside the body of the text and "Name & Address" to indicate a general reference source to a corporate body.

- **Serial Citation:** To the field Serial Citation, other GreyNet publications were likewise understood e.g. citations to the GreyNet Newsletter, Webpage, Notebook, etc.
- **Type of Publication:** Since there was no authority file linked to the field Type of Publication, post-standardization was applied. For example, when the term Website was first used, it was later replaced by the term Webpage. And when the term Article was initially entered the term Journal Article replaced it. In a recent paper by Helmut Artus (Artus, 2004), where he examines grey and commercial journals in the social sciences, the term journal article applies to both. Prior to this study, and in line with a policy maintained by EAGLE, journal articles were not included in the SIGLE database. This insinuated that journal articles were not a type of grey literature, when in fact we will see that today they are among the top five types of grey literature cited.
- **Kind of Publication:** In the field Kind of Publication, a significant number of "Uncertains" as to whether a cited document was grey or not, would later be resolved through literature search, team discussion, and further explanation of the term "commercial publishers", which is traced to the Luxembourg definition of grey literature, 1997. For our purposes here, publishers whose primary or sole activity is publishing for profit, are distinguished from "grey publishers", whose primary activity does not lie in the publishing industry and is not primarily commercial.
- **Self-Citation:** To the field Self Citation, citations not only to oneself but also to ones organisation were added.
- **Name of Cited Source (1st, 2nd, 3rd, 4th):** And finally, to the field(s) Name of (1st, 2nd, 3rd, 4th) Cited Source, if the data entry pertained to a non-person the ISBN rule of thumb was initially used, and the "Title of the work" was first entered. Later it was evident that for GL records, the "Acronym of the Corporate Author" should have taken precedence and was later implemented.

PROJECT LOG: INFORMAL DOCUMENTATION OF THE RESEARCH PROJECT

Throughout all three stages of the research project and perhaps most importantly in the first stage of data collection and processing, the instrument of a project log was implemented. Since four project members would be working in separate offices in two different countries, maintaining a project log in which the questions, comments, and problems encountered could best be captured and recorded. One member of the team was responsible for maintaining the project log, and it was first utilized at the June meeting in Amsterdam where all four-project workers were present. In fact the project log served as an agenda for the daylong meeting as well as a means for the further division of labour among the team members during the remaining two stages of the research project. By the second and last meeting in October, the emphasis on the Project Log was replaced by attention to the Rough Draft of the paper.

Literature Search and Review

SELECTION AND IDENTIFICATION OF THE SOURCES

The second stage in the research project was a literature search and review on citation analysis previously carried out in the field of grey literature and to some extent beyond. The references and sources cited at the end of this paper are divided under three headings, and as such enable the reader to understand their place and emphasis in this research project. Under the first heading are references dealing both with the topic of grey literature and the use of citation analysis. All but two of the twenty-four bibliographic references under this heading are found in the 4th edition of the Annotated Bibliography on the Topic of Grey Literature. It is interesting to note that 10 of these 24 references are related to papers in the GL-Conference Series and that two of the remaining 14 references mention that the original text was written as a paper in the GL-Conference Series. Hence, half of the references under this heading are directly related to papers in the GL-Conference Series. Under the second heading, are the references that deal with citation analysis but which are not specific to the topic of grey literature. And, under the third heading are the sources from which our data for this citation analysis were collected. These comprise the bibliographic references to the first five proceedings published in the GL-Conference Series from 1994 to 2004 and to the 4th edition of the Annotated Bibliography on the Topic of Grey Literature published in 2000.

REVIEW OF THE FINDINGS FROM THE LITERATURE

Examining the issues dealt with in our review of the literature, we find that the use of citation analysis in the study of grey literature has been applied in many disciplines, across disciplines, and in all of the branches of science i.e. the physical, the social, and the humanities. At least 14 specific fields of study are identified. The fields of agriculture and physics are dealt with in multiple studies, while the fields of transportation, social work, environmental protection, education, astronomy, aerospace are each dealt with in one study. When looking at the types of documents that were examined, we find that reports, grey journals, and conference proceedings are accounted for in most of the studies, while e-prints, preprints, and synopses were each the object of only one particular study. An interesting point is that some of the studies that were non-specific to grey literature, also dealt with specific types of grey literature such as theses, dissertations, and grey journals. The studies that were specific to grey literature were the ones most concerned with the fields in a citation record as well as the format of the

citation. The type of grey literature cited, the producers of the document, language of the document, and the publication date were the fields most discussed. Language was the field, which preoccupied the researchers most. English language publications were by far in the majority and local languages such as Finnish were considered barriers to information transfer. Two studies dealing with the life of a citation showed that the highest percentage of citations was within 5 years for journals and up to 10 years for reports. Incomplete citations and errors in citations affect about 12% of the total citations in studies. The impact factor that grey and non-grey citations have on a field of study is of interest to researchers. One study refers to the law of scattering, where 80% of the citations are shown to come from 20% of the journals cited. Stated otherwise, the top 20% of journal titles generates 80% of use. Another study proposes a formula to arrive at **(x)** the intensity of grey literature citing by dividing **(z)** the frequency/proportion of GL references out of all the references cited by **(y)** the frequency/proportion of the articles with GL references out of all the articles examined. The formula would then read $z \div y = x$. And when applied to the results from our research, the intensity of grey literature citing is 59%.

A number of studies, most of which were not specifically dealing with grey literature call for convergence or aggregates of data, where not only citations are used to explain or demonstrate the impact factor but also a number of other empirical data are used. One such study calls for a web invocation portfolio (WIP), where web links (inlinks), standard citations (non-links), bibliographic references, log files, press cuttings, etc., have to be examined together in order to arrive at the true impact factor. It is the role of organisations such as library associations and grey literature producing bodies' i.e. corporate authors to promote and support this type of research in order to demonstrate its uses and applications. These could be to access trends, arrive at a representation/snapshot of a particular field of study, or to influence information policy. In particular, we have seen some studies carried out with grey literature, where citation analysis was used to determine and evaluate appropriate publication channels and others for purposes of collection development. A few studies looked at citation use in order to determine communication across various fields of study. And a couple of studies in particular examined grey literature's implications for scientific communication - differentiating between scientific and non-scientific grey. No doubt, abstracting and indexing services such as ISI Web of science are an invaluable means for carrying out citation analyses making use of automated technologies on an economics of scale. However, grey literature can only really profit from this when they have established a record format, one that is collectively applied. This can be done with the support of those library associations and grey literature producers mentioned earlier. Our review of the literature reveals a number of plans of action and specific recommendations for grey literature - the upshot of various researches. These range from the establishment of national and regional networks to the formulation of guidelines, and the development of course training.

Content and Data Analysis

ANALYSIS OF THE CITATION DATA

The Citation database, which was compiled during this research project, contains 1374 records. 30 of these records (0.02%) are to papers in the GL-Series that do not include references. Each record contains 19 fields, which amounts to 26106 possible entries in the database. This includes fields left blank. The data was extracted from the 139 papers in the first five volumes of the GL-Conference Proceedings. When the 30 papers that did not include references (21%) is subtracted, our citation analysis is then to 109 conference papers. A total of 152 authors/co-authors were responsible for the 139 conference papers. Of these, 108 (71%) are sole or first authors of the papers.

In total 1721 authors/corporate authors (i.e. the sources) were cited in the 1344 records, which was compiled from those papers containing references. When multiple citing is accounted for (i.e. authors cited 1 > times), the number of distinct or unique authors cited in this study is then 1128. There were 72 different types of documents cited of which journal articles and conference papers accounted for the first and second most cited types of documents. There was an average of 12.3 cited references per conference paper. And, looking at the age of the citations, 58% are <4 years old.

When standard bibliographic citations are compared with hyperlinked citations, we find that 51% of the standard citations are to grey literature, 48% to commercial publications, and 1% is uncertain. Persons accounted for 84% of the standard citations, while non-persons or corporate authors accounted for 16%. If we compare this with the hyperlinked citations, we find that 90% are to grey literature, 5% to commercial publications, and 5% is uncertain. Persons accounted for only 26% of the hyperlinked citations, while non-persons or corporate authors accounted for the other 74%.

If we look at self-citations, which in this study also include citations to one's own organization, this accounts for 10% of the citations. In this category, 26% is non-applicable, because the citations are to non-persons or corporate authors with no affiliation to the citing author. If we turn to serial citations, that is a citation to one of the earlier papers in the GL-Conference Proceedings or to one of GreyNet's other publications (e.g. Annotated Bibliography, Guidebook, Notebook, etc.), nearly 8% (0.0775) fall under this heading. Citations in the first volume in the Conference Series (GL'93) would then by our definition of serial be excluded here. The cited documents were in 13 different languages in which English accounted for 84.7% followed by Italian 5.59%, French 2.41%, Spanish 1.86%, German 1.32%, and Dutch 1.24%. The remaining 7 languages (Indian, Indonesian, Japanese, Korean, Polish, Portuguese, and Russian) each

accounted for less than 1% of the cited documents. If we compare this with the language of the Citing Authors, we find that 65% of the GL Conference Papers are by non-native English speakers. Of the 72 types of publications cited, an average of the top five over the conference series show Journal Articles at 47%, Conference Papers at 19%, Books/Monographs at 17%, Reports at 14%, and Conference Proceedings at 3%. These percentages are based on standard citations; however, when hyperlinked citations are included, then the Webpage ousts the Conference Proceedings from the fifth place. The last conference in the series (GL5) follows these averages almost entirely. There are only 17 types of commercial publications cited, while 72 types of grey literature publications are cited. Reports, which held second place at the first conference (GL93) dropped to third place by the second conference (GL95) and to fourth place by the third conference (GL97). These had traditionally been the mainstay of GL, but this research shows that conference papers have taken and hold second place.

A COMPARISON WITH OTHER DATA

When we compare the authors cited in the conference proceedings with the authors cited in the index to the 4th edition (2000) of the Annotated Bibliography on the Topic of Grey Literature, we find an average match of 24.7%. Some 630 authors are indexed in this Annotated Bibliography, and it contains over 750 bibliographic references.

Cited Authors in the Proceedings compared with Authors indexed in the Bibliography

1 st Cited Source	2 nd Cited Source	3 rd Cited Source	4 th Cited Source
330/1344 = 24.6%	87/293 = 29.7%	21/89 = 23,6%	8/38 = 21%

The authors indexed in the Annotated Bibliography are authors who specifically deal with grey literature in their research and writing i.e. the GL meta-authors. It might then be safe to speculate that this quarter of the authors cited in the conference proceedings could be matched by another quarter that deals with grey literature, but does not explicitly adhere to the term. The remaining half of the authors cited in the proceedings may then have intended no reference to grey literature at all. In short, half of the citations in our primary analysis are directly or indirectly related to grey literature and the other half are non-related. If we were to rely on the type of documents cited as an indicator, then the ratio would not be 50-50 but instead would increase to nearly 60-40 with grey citations outnumbering commercial citations by 790 to 534 with still 20 uncertain. Either way, this is a very healthy position for grey literature within the entire field of information. It offers foundation and links grounded in broader theory and practice, while it demonstrates a clear and identifiable entity within the whole. If the meta-authors in grey literature would have remained predominately or solely within the research and findings of those with the same signature - then over the past quarter of a century, grey literature would have withered and become dormant. And, this is certainly not the case.

FINDINGS OF THE RESEARCH

Results from our research project indicate that many of the problems encountered in the citation analysis of books and journal articles are also found in the citation analysis of grey literature e.g. typographical errors, omissions, and inconsistencies. These problems are augmented in a citation analysis of grey literature, because it is necessary to indicate the "type of publication" (i.e. report, conference paper, Webpage, etc.) as well as the "kind of publication" (grey or commercial). The more complete and accurate the citation, the easier it is to make these determinations. Hyperlink citations prove to be even more difficult for the project workers than standard bibliographic ones. Citation analysis of grey literature requires not only the name(s) of the individual author(s), but also the corporate author or organisation responsible for the publication.

Across disciplines, reports, grey journals, and conference proceedings are the main sources used in research and authorship. This finding supports the premise that the GL-conference proceedings are of significant importance to the field of grey literature. This field of information studies currently does not have a flagship journal. In 2000, this was attempted with the launch of IJGL (International Journal on Grey Literature). But, it only lasted one year with the publication of one volume containing four issues. Interestingly, a third of the articles in this volume first appeared as conference papers in the GL-Series. De facto, the GL Conference Proceedings has taken on a flagship role. And it is not an unlikely one, because as Buckley purports (Buckley, 1997), conference papers in proceedings are similar to articles in journals. Another attempt in the past was to bring together a selection of the GL papers for publication in a special issue of a journal, first published in PRQ (Publishing Research Quarterly volume 13, no.2, 1997). With the re-launch of the GL-Series in 2003, this has now led to a cooperative publishing agreement with PRQ, where annually an issue of this quarterly journal focuses on the topic of grey literature. Collection holdings of the GL conference proceedings and papers should be comprehensive and available from more than one centre or clearinghouse. Currently, the British Library Document Supply Centre in Boston Spa is to our knowledge the only dedicated centre for this purpose. And, this arrangement between the GL Proceedings publisher, TextRelease, and the British Library has been recently reaffirmed. Contacts with other national libraries and centre are likewise being pursued, for example with the Netherlands Depot and INIST in France.

Where once reports were the main type of grey literature, we see that grey journals and conference papers are the most heavily cited types of grey literature. Since the onset of the GL Conference Series, more than a decade ago, English has been the language of publication. This has enabled the authors' research findings to reach global communities. We see that only about 35% of the papers in the GL-Conference series are from authors from English Speaking countries, while non-native English speakers account for 65% of the content contributions. One of the drawbacks of the GL Proceedings brought out in this study indicates the lack of standardization applied to citations. This aspect is one in which both the authors and their works would most readily stand to gain were a standard applied. One which would capture and accommodate the various types of grey literature and which would be in line with standards applied by existing abstracting and indexing services such as ISI, LISA/CSA, etc.

From our research findings, we have compiled the following guidelines for grey literature citations.

GUIDELINES FOR GREY LITERATURE CITATIONS	
General	
i.	All conference papers should contain references
ii.	Standardization should be maintained among the citations provided
iii.	The more complete and accurate a citation, the more guarantee of a paper's content and subsequent review.
	<i>Examples of Errors that were encountered and corrected</i>
	Hedlund, G. (GL97); Hedlung, G. (GL99) – (GL97 is correct)
	Hugenholtz, <u>B.P.</u> (GL95); Hugenholtz, <u>P.B.</u> (GL99 – (GL99 is correct)
	Rozkuszka, <u>W.</u> (GL95); Rozkuszka, <u>D.W.</u> (GL99) – (Rozkuszka, <u>W.D.</u> neither was complete or correct)
Specific	
iv.	Endnotes are preferred and should be numbered
v.	Hyperlinks need the accompanying name of resource and date; a simple URL is not acceptable
vi.	If the citation is to a corporate author, the acronym takes precedence
vii.	If the document type is known, it should be stated at the close of a citation.
viii.	If a citation is revised and refers to an edited and/or abridged work, the original source should also be mentioned.

While these guidelines do as yet offer a citation style for grey literature; hopefully, this will be a spin-off from our research. Perhaps, starting with the 7th volume of the GL-Conference Proceedings, a standard or citation style will be in place. Likewise, a renewed contact with LISA, is underway that will hopefully lead to re-establishing and exchange agreement, where the abstracts of the GL Conference Proceedings would be given to LISA and in turn LISA would held facilitate the research and work carried out by GreyNet, Grey Literature Network Service. Clearly, we see that the impact of our study is leading GreyNet to establish and formalize a policy, which would have lasting benefit for authors and researchers both within and outside the GL-conference series. A policy which would enhance the published proceedings, bring excellence outside of the grey circuit through the cooperative publishing agreement with PRQ, guarantee document delivery and/or loan of the full-text of all of the conference papers in the series via an ongoing agreement with the British Library and hopefully other depots and clearinghouses of longstanding. Yet another impact that our citation analysis could have is on the advancement of grey literature as a field of information study in its own right. The core papers could be compiled and published as a reader(s) used by colleges and universities in course modules on grey literature. Also, Auger's first roadmap of GL systems and services, which reached its 4th and last edition in 1998 (Auger, 1998) could benefit by way of a thorough and comprehensive revision in light of the research and findings from the GL-Conference Proceedings.

Conclusions and Further Recommendations

The concern at the outset of this research project was to apply the instrument of citation analysis to the work of authors in the GL-Conference series, as well as authors whom they cited both in grey and commercial publishing. A further concern was to determine the value of the GL Conference Proceedings for research in the field of grey literature by treating it as a type of grey literature and comparing it with other types. Our findings indicate that an impact in one of the three areas is connected in some way to the other two. We find that the term "commercial" in the Luxembourg definition of grey literature needs to be further elaborated in order to better serve its purpose. The meta-authors should not underestimate

the explicit use of the term grey literature in the title of their works. Further, based on this research, a set of guidelines has been compiled for citations used in grey literature. These will be implemented starting with the 7th volume of the GL-Conference Proceedings; however, it should not yet be taken as a standard or citation style for grey literature. The authors of this paper would hope that by maintaining the Citation Database and making it available to researchers for secondary analysis, a follow-up study would be possible. Perhaps this could be combined with a research proposal to ISI/ASIS, who offers an annual grant for research dealing with citation analysis.

In close, GreyNet is interested in carrying out a survey among the authors and co-authors in the GL Conference Series, including the 40 new contributors to GL6. It is believed that this would provide evidence of a wider impact for research than can be shown by citation analysis alone. Finally, GreyNet would like to explore the possibility of launching a grey journal, where better coverage and exposure of research in the field of grey literature is achieved not just for the portion that crosses over to commercial journals. As this study indicates, journal articles and conference papers are the top two types of cited publications both in grey and commercial publishing. Would this then not be considered a mandate in itself?

References and Sources

REFERENCES SPECIFIC TO GREY LITERATURE AND CITATION ANALYSIS

1. Aina, L.O. (1987), Grey literature and agricultural research in Nigeria. - In.: IAALD Quarterly Bulletin. - Vol. 32, no. 1 (1987), pp. 47-50.
2. Aina, L.O. (1996), Promoting the exploitation of grey literature in Africa. - In: GL'95 Conference Proceedings: The Second International Conference on Grey Literature "Grey Exploitations in the 21st Century" held in Washington D.C., 2-3 November 1995 at the Catholic University of America. - Amsterdam : GreyNet/TransAtlantic, 1996, pp. 63-73. - (GL-conference series, ISSN 1385-2316 ; No. 2). - ISBN 90-74854-08-7
3. Aina, L.O. (1998), The management of grey literature as a component of a library and information science curriculum. - In: GL'97 Conference Proceedings: The Third International Conference on Grey Literature "Perspectives on the Design and Transfer of Scientific and Technical Information" held in Luxembourg, 13-14 November 1997 at the Jean Monnet Building. - Amsterdam : GreyNet/TransAtlantic, 1998, pp. 251-258. - (GL-conference series, ISSN 1385-2316 ; No. 3). - ISBN 90-74854-17-6
4. Alberani, V. and P. De Castro Pietrangeli (1994a), The initiatives of a library association for the promotion of grey literature. - In: GL'93 Conference Proceedings : The First International Conference on Grey Literature "Weinberg Report 2000" held in Amsterdam, The Netherlands, 13-15 December 1993 at the RAI Congress Centre. - Amsterdam : GreyNet/TransAtlantic, 1994, pp. 50-55. - (GL-conference series, ISSN 1385-2316 ; No. 1). - ISBN 90-74854-03-6
5. Alberani, V. and P. De Castro Pietrangeli (1994b), Grey literature in information science : production, circulation, and use. - In: GL'93 Conference Proceedings: The First International Conference on Grey Literature "Weinberg Report 2000" held in Amsterdam, The Netherlands, 13-15 December 1993 at the RAI Congress Centre. - Amsterdam : GreyNet/TransAtlantic, 1994, pp. 56-63. - (GL-conference series, ISSN 1385-2316 ; No. 1). - ISBN 90-74854-03-6
6. Alberani, V. and P. De Castro Pietrangeli (1995), Grey literature in information science : production, circulation and use. - In: Inspel 29 (4), pp. 240-249. - ISSN 0019-0217.
7. Berman, Y. and A.S. Eaglstein (1994), The knowledge base of social work : a citation analysis. - In: Aslib Proceedings, 46 (9) Sep 94, pp. 225-230. - ISSN 0001253X
8. Cordes, R. (2002), Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an international marine scientific advisory body. - In: Canadian Journal of Information and Library Science. 2002, 27 (3), pp. 109-127.
9. Di Cesare, R. (1994a), The evaluation of GL impact in Physical sciences using bibliometric indicators: preliminary results. - In: GL'93 Conference Proceedings: The First International Conference on Grey Literature "Weinberg Report 2000" held in Amsterdam, The Netherlands, 13-15 December 1993 at the RAI Congress Centre. - Amsterdam : GreyNet/TransAtlantic, 1994, pp. 64-76. - (GL-conference series, ISSN 1385-2316 ; No. 1). - ISBN 90-74854-03-6
10. Di Cesare, R. (1994b), The evaluation of grey literature impact using bibliometric indicators : the case of physical sciences. - In: Proceedings of the 18th International Online Information Meeting, London, 6-8 December 1994. Edited by David I. Raitt and Ben Jeapes, Oxford and New Jersey, Learned Information (Europe) Ltd, 1994, pp. 405-414.
11. Di Cesare, R. and Cesare Sala (1996), The use of grey literature in the agricultural economics field: a quantitative analysis. - In: GL'95 Conference Proceedings: The Second International Conference on Grey Literature "Grey Exploitations in the 21st Century" held in Washington D.C., 2-3 November 1995 at the Catholic University of America. - Amsterdam : GreyNet/TransAtlantic, 1996, pp. 157-168. - (GL-conference series, ISSN 1385-2316 ; No. 2). - ISBN 90-74854-08-7

12. Heij, D.G. van de, J. van der Burg, I.R.C. Cressie and M. Wedel (1990), Comparative analysis of the penetrative capacity of synopses and of full papers unrelated to the synopses published in the same broad-scope journal. - In: Journal of Information Science. - Vol. 16 (1990), pp. 155-164.
13. Kortelainen, T. (1993), Characteristics of internationally successful peripheral periodicals. - In: IATUL Proceedings (New Series), 2, pp. 110-122. - ISSN 0966-4769
14. MacDonald, B.H., R.E. Cordes, and P.G. Wells (2004), Grey Literature in the Life of GESAMP, an International Marine Scientific Advisory Body. - In: GL5 Conference Proceedings : The Fifth International Conference on Grey Literature "Grey Matters in the World of Networked Information" 4-5 December 2003. - Amsterdam : GreyNet/TextRelease, 2004, pp. 50-63. - (GL-conference series, ISSN 1385-2316 ; No. 5). ISBN 90-77484-01-9.
15. McKiernan, G. (2000), ResearchIndex : autonomous citation indexing on the Web. - In: International Journal on Grey Literature, IJGL. - MCB University Press. Vol. 1, No. 1, 2000, pp. 41-46. - ISSN 1466-6189
16. Osaniyi, L. (1996), IITA conducts study on the use of grey literature. - In: Quarterly Bulletin of the International Association of Agricultural Information Specialists 41 (3/4) 1996, pp. 263-264. - ISSN 1019-9926
17. Osif, B.A. (2000), International Transportation Literature: An Analysis of Citation Patterns, Availability and Research Implications for the Transportation Community. - In: GL'99 Conference Proceedings : The Fourth International Conference on Grey Literature "New Frontiers in Grey Literature" held in Washington D.C., 4-5 October 1999 at the Kellogg Conference Center. - Amsterdam : GreyNet/MCB UP, 2000, pp. 47-60. - (GL-conference series, ISSN 1385-2316 ; No. 4). - ISBN 9074854-24-9
18. Patil, Y.M. and P.S.G. Kumar (1997), Grey literature in Pedology : how grey is it? - In: International Information Communication & Education, 16 (2) Sep 1997, pp. 171-181. ISSN 0970-1850
19. Patil, Y.M. and P.S.G. Kumar (1998), Pedological literature analysis. - In: Library Science with a Slant to Documentation and Information Studies, 35 (2) June 1998, pp. 125-132. - ISSN 0254-2553
20. Pinelli, T.E. and V.L. Golich (1998), The Reader Preferences of NASA Technical Reports in Electronic Format: Results of a Survey. - In: GL'97 Conference Proceedings : The Third International Conference on Grey Literature "Perspectives on the Design and Transfer of Scientific and Technical Information" held in Luxembourg, 13-14 November 1997 at the Jean Monnet Building. - Amsterdam : GreyNet/TransAtlantic, 1998, pp. 46-72. - (GL-conference series, ISSN 1385-2316 ; No. 3). - ISBN 90-74854-17-6
21. Ramadevi, T. and A. Takalkar (2000), Use Pattern of Grey Literature in Rural Development : An Indian case study. - In: GL'99 Conference Proceedings : The Fourth International Conference on Grey Literature "New Frontiers in Grey Literature" held in Washington D.C., 4-5 October 1999 at the Kellogg Conference Center. - Amsterdam : GreyNet/MCB UP, 2000, pp. 260-275. - (GL-conference series, ISSN 1385-2316 ; No. 4). - ISBN 90-74854-24-9
22. Vinogradov, A.N. and L.M. Kabdulova (1995), Grey literature as a main source of information about modern Russian knowledge on the Arctic Region development. - In: Proceedings of the 15th Polar Libraries Colloquy, Bipolar information initiatives : the needs of polar research. Cambridge, 3-8 July 1994. Edited by D.W.H. Walton, W. Mills and C.M. Phillips. Bluntisham, UK, Bluntisham Books for Polar Libraries Colloquy Steering Committee, 1995, pp. 98-100.
23. Youngen, G.W. (1998), Citation patterns to traditional and electronic preprints in the published literature. - In: College & Research Libraries, 59 (5) Sep 1998, pp. 448-456. ISSN 0010-0870

REFERENCES SPECIFIC TO CITATION ANALYSIS, NON-SPECIFIC TO GREY LITERATURE

24. Buckley, C.E. (1997), Core conservation biology monographs: A citation analysis. - In: Reference Services Review, Fall/Winter 1997, pp. 147-159
25. Kwak, M. (2002), Using Bibliometric Journal Citation Analysis as a Technique to Assess Trends in School Psychology Journal Publications Between 1995-1999. - Doctoral Dissertation 2002, Western Michigan University. - <http://www.wmich.edu/grad/dissertation/dis-archive/kwak.html>
26. Leydesdorff, L. and O. Amsterdamska (1990), Dimensions of Citation Analysis. - In: *Science, Technology and Human Values* 15 (1990) 305-335. - <http://users.fmg.uva.nl/lleydesdorff/sthv90/>
27. Okiy, R.B. (2003), A citation analysis of education dissertations at the Delta State University, Abraka, Nigeria. - In: Collection Building, Volume 22, Number 4, 2003, pp. 158-161
28. Sylvia, M.J. (1998), Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies. - In: Collection Building, Volume 17 · Number 1 · 1998 · pp. 20-28, ISSN 0160-4953
29. Thelwall, M. (2002), Research dissemination and invocation on the Web. - In: Online Information Review, Volume 26, Number 6, 2002, pp. 413-420

SOURCES USED FOR THE DATA COLLECTION

30. Farace, D.J. (1994), GL'93 Conference Proceedings : First International Conference on Grey Literature, Amsterdam RAI, December 13-15, 1993. - Grey Literature Network Service. - Amsterdam : GreyNet/TransAtlantic, February 1994. - XIV, 418 p. ; 30 cm. - Index. - (GL-conference series, Conference proceedings, ISSN 1386-2316 ; No. 1). - ISBN 90-74854-03-6
31. Farace, D.J. (1996), GL'95 Conference Proceedings : Second International Conference on Grey Literature : Grey Exploitations in the 21st Century, Washington D.C., 2-3 November 1995. - Grey Literature Network Service. - Amsterdam : GreyNet/TransAtlantic, February 1996. - X, 266 p. : ill. ; 30 cm - Indexes. - (GL-conference series, Conference proceedings, ISSN 1386-2316 ; No. 2). - ISBN 90-74854-08-7
32. Farace, D.J. and J. Frantzen (1998), GL'97 Conference Proceedings : Third International Conference on Grey Literature : Perspectives on the design and transfer of scientific and technical information, Luxembourg DGXIII, November 13-14, 1997. - Grey Literature Network Service. - Amsterdam : GreyNet/TransAtlantic, March 1998. X, 294 p. ; 30 cm. - (GL-conference series, Conference proceedings, ISSN 1386-2316 ; No. 3). - ISBN 90-74854-17-6
33. Farace, D.J. and J. Frantzen (2000), GL'99 Conference Proceedings : Fourth International Conference on Grey Literature : New Frontiers in Grey Literature, Washington, D.C., 4-5 October 1999. - Grey Literature Network Service. - Amsterdam : GreyNet/MCB UP, January 2000. - ill. XII, 322 p. ; 30 cm. - Author Index. - (GL-conference series, ISSN 1386-2316; No. 4). - ISBN 90-74854-24-9
34. Farace, D.J. and J. Frantzen (2000), Annotated Bibliography on the Topic of Grey Literature : A Public Enterprise in Editing and Review. - Grey Literature Network Service. - Amsterdam : GreyNet/MCB UP, May 2000. - 4th Edition. - ISBN 90-74854-26-5
35. Farace, D.J. and J. Frantzen (2004), GL5 Conference Proceedings : Fifth International Conference on Grey Literature : Grey Matters in the World of Networked Information, 4-5 December 2003. - Grey Literature Network Service. - Amsterdam : GreyNet/TextRelease, January 2004. - 162 p. - (GL-conference series, ISSN 1385-2316 ; No. 5). - ISBN 90-77484-01-9

Author Information

Dr. Joachim Schöpfel, Responsable du departement
Resources documentaires, INIST-CNRS
Institut de l'Information Scientifique et Technique
Centre National de Recherche Scientifique
2, allée du Parc de Brabois
54514 Vandoeuvre-Les-Nancy Cedex
France

EML: schopfel@inist.fr
URL: <http://www.inist.fr/>

Grey Literature in Energy: 5 Years Later

Deborah E. Cutler, U.S. Department of Energy (DOE)
Office of Scientific and Technical Information (OSTI)

Abstract

At GL'99, the Department of Energy's Office of Scientific and Technical Information (DOE OSTI) presented a paper titled "Grey Literature in Energy: a Shifting Paradigm." Five years later, the paradigm continues to shift, less radically than the change from paper and microfiche to electronic, but with significant benefit to users worldwide. OSTI's efforts continue this shift, moving even further away from the connotation of 'grey' as meaning 'hard to get' literature. Along with its domestic and international partners dealing with science and energy information, OSTI serves as a leader in tools for accessing grey literature in addition to its role as a major repository. The paper will highlight the collections, tools, and partnerships that have allowed OSTI to truly bring science and energy information to the computer desktop.

OSTI first and foremost partners within DOE. Energy-related grey literature residing at OSTI, along with the grey literature residing at other locations throughout the DOE complex of laboratories, is made available and searchable to the public through OSTI IT systems. While this is not new, several related developments are, including OCR steps for image data, remote indexing, and single page searching. Recently, special initiatives with Google and Yahoo/MSN in the past year have resulted in users now being able to find DOE's grey literature residing within OSTI databases when using these browsers. Providing yet a further avenue, OSTI's latest initiative involves the possibility of joining CrossRef, meaning that Digital Object Identifiers (DOI's) traditionally associated with published literature may soon be assigned to DOE's grey literature collection.

Partnerships with other U.S. federal agencies have resulted in a massive collection of science-related grey literature being available through a single interface. Offered to users initially through the GrayLIT Network developed by OSTI, this has now been broadened further through the e-government success story known as Science.gov. 'Deep Web' searching allows users to go beyond the typical browser in searching web pages, down to the database and full text levels and then across U.S. government agencies.

Partnerships with the International Energy Agency's Energy Technology Data Exchange (ETDE) and the International Atomic Energy Agency's International Nuclear Information System (INIS) have augmented the domestic collection into a major global resource known as ETDEWEB, managed and operated by OSTI, and the INISDB online, managed by INIS. Current OSTI and INIS efforts to convert the older microfiche collections of grey literature into electronic form will serve to advance the knowledge management and preservation initiatives of these major repositories.

It is almost hard to imagine that less than 10 years ago users requesting U.S. government reports often had to wait many days and sometimes weeks to receive their paper copies or microfiche, usually via what we today call 'snail mail.' Five years ago, at GL'99, the Department of Energy's Office of Scientific and Technical Information (DOE/OSTI) presented a paper titled "Grey Literature in Energy: a Shifting Paradigm." The paper described the evolution of DOE reports delivery by OSTI from the paper and microfiche era into the electronic age of the Internet. The late 1990's saw the introduction of OSTI's DOE Information Bridge, an innovative web system that allowed users to find DOE reports and view or download the full text. In many ways, this easy access took the 'hard to get' part of the grey literature definition out of the picture. Five years after presenting that paper, OSTI remains on the forefront of technology and techniques used in making grey literature even more easily accessible in a greater variety of ways. For instance, in just the past year, thanks to two OSTI initiatives, Internet searchers using the search engines Yahoo! or Google can now find individual DOE reports showing up on hit lists. This greatly expands the typical Web surfer's ability to find DOE grey literature into the depths of the Deep web. This paper will highlight more details on these and other initiatives plus give a current state of the collections, tools, access points, and partnerships that have allowed OSTI to truly bring science and energy information to the electronic desktop.

First, perhaps a little background on DOE and OSTI is in order. As one of the largest sponsors of research and development for the U.S., DOE supports thousands of research projects annually at the many DOE laboratories and universities across the country through contracts and grants. Researchers are required to document their research results in reports and can also submit articles to journals and present papers at conferences. OSTI partners with the DOE sites to serve as DOE's central repository for DOE research results. OSTI's mission is to advance science and sustain technological creativity by making these R&D findings available and useful to DOE researchers and the American people. Through OSTI's collections and services, the DOE research results are made available to scientists, researchers and engineers in the DOE community as well as to academia, the international science community, and science-attentive citizens. In addition to DOE-funded research, OSTI also provides research information of interest to DOE through partnerships with other federal agencies and international communities.

DOE Partnership Initiatives

Historically, OSTI offered paper publications that documented R&D results. Beginning in the 1970's, research results were announced in electronic bibliographic databases that included metadata (such as titles, authors, research organization, date of publication, subject indexing and abstracts) to enable searching and then provided an availability notice for access to the full text of the reports. Public access to the reports was usually handled through the U.S. Department of Commerce's National Technical Information Service (NTIS) or the Government Printing Office's (GPO)'s Depository Library system. Paper and microfiche were the public's only options for full text delivery up through the mid-1990's. As mentioned in the introduction, document delivery took days to weeks.

DOE Information Bridge

In the mid- to late-1990's, OSTI focused its efforts on electronic full text delivery of reports, cutting delivery time to minutes. Initially a DOE-only system, OSTI soon partnered with GPO to offer the public this direct access via the DOE Information Bridge (<http://www.doe.gov/bridge>). The number of reports available in September 1999 was 43,000. Today, there are 85,000 DOE reports available for viewing and downloading through this system.

Users of the Information Bridge system have the option to search the bibliographic information, or they can also search the full text for much of it, dependent on format. New developments for the Information Bridge system since the 1999 paper include the following. Regarding coverage, the system now includes reports dating from 1995 forward, with the 1994 reports due to be available within the next six months. Regarding the format of the full text, PDF files with searchable full text have replaced TIFF images as the standard for most documents, allowing additional search capabilities. Regarding the metadata format, a Dublin Core-like format is used as the base format, with displays and search indexes changed accordingly. Full text indexing is now also done on documents residing on laboratory sites as long as they are in indexable formats. And lastly, single page searching has been revised to also use PDF pages rather than the TIFF images.

Google and Yahoo!

OSTI recognized that the main difficulty with accessing the DOE report literature was that you had to be in Deep web database systems such as the Information Bridge to find it. Yes, DOE's literature was accessible, but only if the user knew about the system. In today's terminology, this illustrates the difference between the Deep web (typically information that resides below the surface in databases) and the surface web, which reflects the main levels of web pages. Search engines normally only search the surface web. Given that the volume of information in the Deep web could be over 500 times greater than the surface web (according to figures referenced in many places on the web) it is no wonder that the major search engines index this information with discrimination, seeking to include those with high quality information or those willing to pay for such a service. Thus, the situation coming into the year was that typical users would NOT find many DOE reports when entering an Internet search in such search engines as Yahoo! or Google.

As mentioned briefly in the introduction, two of the exciting new partnerships OSTI has made during the past year have led to this situation changing. Users searching Google and Yahoo! now DO get hits on DOE's grey literature. Once found, the user can then access OSTI's systems to see the metadata for the record, and then directly go to the full text of the report, if available. The methodologies used by each of the search engines differ in the way they wanted to interact with OSTI based on their system designs. Both involve OSTI making detailed information available for the search engines to find and then index. The way the Google system interacts is that OSTI makes an enhanced site map available which includes groupings of the database information by date, referencing titles and links to bibliographic pages, which are also available for indexing. Google then periodically 'crawls' these data to add to its indexes. The Yahoo! methodology, on the other hand, works differently. OSTI has created an XML Web service to provide database information, which is then captured by Yahoo! for monthly indexing. The difference seems to be that Yahoo! retrieves the information all at one time while the Google system retrieves portions at a time, more slowly. OSTI's observations over the past year have shown differences in what all gets indexed, but the net result is that users of either search engine now have access to a significant amount of DOE's grey literature.

Yahoo! also offers web metrics to the system administrators, allowing OSTI to track usage. According to David Mandelbrot, vice president, Yahoo! Search Content, "*Yahoo! Search is committed to creating the best search experience on the Web. We are excited about our direct relationship with OSTI as it further enables us to provide users a more comprehensive search experience with specialized content that is otherwise hard to find on the Web.*"

Energy Citations Database

One of the other new OSTI public databases offered since the 1999 paper is the Energy Citations database (<http://www.osti.gov/ECD>). The focus of this database is metadata reaching back to the 1940's (both grey literature and non-grey literature are included) with links to the full text of reports for the same segment of records found in the Information Bridge. OSTI has also arranged to have Google and Yahoo! index records from this database.

As a result of all of these access points to DOE's grey literature collection, the average number of downloads annually totals 400,000 documents.

DOI initiative

Digital Object Identifiers (DOIs) are becoming very common and useful in allowing users to link to the full text of journal articles, conference information and more. The links normally take the user to the publishers' sites where, if their organization already has a subscription, the full text can be viewed free of charge. If no subscription is in place, most publishers offer a pay per view option, and some are even free. Along with its international partners (more on those partnerships later in the paper) OSTI has been actively exploring making the DOIs part of the metadata records, thus increasing the percentage of records where users have a direct path to the full text at the desktop. OSTI is also looking at the feasibility of having DOIs assigned for the DOE report collection. Developments in this area are in the negotiation phase.

Other Federal Agency Partnership Initiatives*GrayLIT Network*

OSTI also has separate Deep web-based initiatives that allow users to search for information contained in Deep web databases across U.S. federal agencies. Also referred to by OSTI as a distributed searching application, one of the first developments in 2000 was the GrayLIT Network. This system offered the capability to search the grey literature collections from four federal agencies: the Department of Defense's Defense Technical Information Center (DTIC), the Environmental Protection Agency (EPA), the National Aeronautics and Space Administration (NASA), and of course, DOE/OSTI.

Science.gov

While the GrayLIT system remains operational, the newer Science.gov initiative involving many more federal agencies and over 47 million pages of information has garnered more interest and support. An interagency working group of 17 scientific and technical information organizations from 12 major science agencies developed Science.gov. OSTI has played a major role in the development of the system, which resides on OSTI systems. OSTI's street address has even been changed to 1 Science.gov Way to help raise visibility of this federal initiative. The distributed searching portion uses something called Explorit, which has recently added some relevancy ranking capabilities. Science.gov debuted in late 2002 and was upgraded in May 2004. It offers Deep and surface web searching of over 30 databases and more than 1,700 science Web sites. While a good deal of the information in these federal systems is NOT grey literature, it is a significant technological achievement worth noting and an example of what agencies can do when they work together toward a common goal. It has received accolades for its contribution to the President's Management Agenda for expanding electronic government services to citizens.

Federal R&D Project Summaries

What has always been some of the 'greyist' literature in the R&D world is related to research in progress – the information that describes what research is planned and underway as opposed to documented research results. Typically, there is no further 'document' available, only the abstract-type description. For some time, OSTI has made available the DOE project summaries, but fairly recently, several agencies came together to offer the Federal R&D Project Summaries (<http://www.osti.gov/fedrnd/>) site. This site provides a portal to information about Federal research projects from six federal agencies' R&D systems:

- DOE R&D Project Summaries Database
- EPA Science Inventory
- National Institutes of Health CRISP (Computer Retrieval of Information on Scientific Projects) Database
- National Science Foundation (NSF) Awards Database
- Small Business Administration Technology Resources Network
- U.S. Department of Agriculture's (USDA) Current Research Information System (CRIS).

DOE/OSTI provides public access to this research tool through GPO Access in partnership with the Government Printing Office. Over half a million summaries are currently available.

International Partnership Initiatives

IEA/ETDE

OSTI's longstanding involvement in the international information exchange has also progressed in the grey literature area since the 1999 paper. Since its inception in 1987, OSTI has served as the Operating Agent for one of the major exchanges known as the Energy Technology Data Exchange (ETDE), an implementing agreement under the International Energy Agency. Currently 16 countries cooperate to exchange energy-related information in all energy areas, including policy. Shortly following the 1999 GreyLit conference, ETDE debuted *ETDEWEB* (ETDE World Energy Base) at IEA Headquarters in Paris. *ETDEWEB* (<http://www.etde.org/ETDEWEB>) was developed by OSTI for ETDE, based on the same concept as the DOE Information Bridge, but it includes information from all participating countries and international partnerships, rather than just US information. *ETDEWEB* currently has close to 3.5 million metadata records dating back to 1974 and over 111,000 electronic full text documents directly available. As with the DOE reports, these documents are also mostly considered grey literature. Close to 50,000 newer records also have links to electronic full text or to websites where the full text can be obtained. Recent initiatives with CrossRef involving DOIs for journal records will see over 250,000 DOI's being made active in the metadata records in November. Although not considered grey literature, as with other initiatives, the goal is to facilitate user access to the source document when at all possible.

ETDEWEB is available to anyone in its member countries with registration, and ETDE's Executive Committee has just recently permitted access to many developing countries as well. Most member countries are able to offer *ETDEWEB* access free of charge. Due to the cost of participation in the agreement and providing input to the database, some member countries do charge a small fee. As the U.S. delegate organization, OSTI offers free access to users in the U.S.

IAEA/INIS

The other major international exchange that OSTI is a part of is the International Atomic Energy Agency's International Nuclear Information System (INIS). OSTI is the U.S. Liaison Officer organization to INIS. INIS (<http://www.iaea.org/inis>) began in 1970 and currently has 112 countries and 19 international organizations involved in exchanging information on the peaceful uses of nuclear energy. INIS offers an Internet database (2.5 million records) of this information from Vienna as well as a CD-ROM product of both the bibliographic information and a separate product for the recent full text. Free access to the Internet database is allowed for universities and contributing organizations in each participating country. Other paid subscriptions are available for individual and corporate users). Coming soon in the next several months will be Internet access to much of the full text (grey literature) cited in the database, dependent on getting country-permissions to post electronically. This has been the top initiative requested by users to facilitate more direct access to this valuable collection of nuclear-related grey literature. Per the partnership between INIS and ETDE, much of the bibliographic information from INIS is included in the *ETDEWEB* database.

In summary, OSTI has been on the forefront of many initiatives to bring users worldwide closer to the full text of both grey and conventional literature coming from many sources -- within DOE, within other federal agencies, and internationally. But this could not have happened without the innovative thinking of OSTI and its partners to develop and implement new technologies. Their willingness to collaborate together and achieve such initiatives in such a short span of time is unprecedented in the government arena. OSTI will continue to strive to build on these partnerships plus add to the collections available. It is hoped by the author that in another five years, there will be even more exciting achievements to share.

Author Information

Ms. Deborah E. Cutler,
International Program Manager
Office of Scientific & Technical Information, OSTI
U.S. Department of Energy, DOE
1 Science.Gov Way
Oak Ridge, TN 37830
USA

EML: CutlerD@osti.gov
URL: <http://www.osti.gov>
<http://www.osti.gov/graylit/index.html>

A study of teachers and researchers practices with digital documents - grey or not

Céline Bourasseau, Cédric Dumas
EMN, Ecole des Mines de Nantes

Abstract

Large-scale consumers of information such as lecturers and researchers have nowadays widely adopted the digital document. These professionals cannot suffice with disaggregate data but instead need full text documents. These documents include the research production of their colleagues and the teaching resources designed both within and outside their institution – be they commercially published or not. The main source of information for these professionals is the Internet, which has become a victim of its own success. If lecturers and researchers claim to have gained better information accessibility, thanks to the Web, they still must account for the time needed to examine the results obtained. Moreover, many of these net-users still camp with difficulties in information retrieval, where all too often their results are unsuccessful or unsatisfactory.

This paper is based on a survey that examines the use of and performance with digital documents by university lecturers and researchers at an Engineering College in Nantes on the West coast of France. The study was conducted during the year 2003 with about 70 persons from various fields and disciplines including physics, computer science, sociology, etc. This study looks at the lecturer-researcher in the capacity of information seeker and reveals a rather sedentary and autonomous figure, one who first relies on the resources offered by his/her own computer. Even though they are partly unsatisfied, they claim no time to waste on improving their information search skills. New tools bringing relevant, rich, and reliable scientific information and documentation are of interest to them. Certainly, if this would help them capture that which would otherwise have been neglected when only classical search techniques are applied. However, these tools must be simple to use, fast, and available where and when needed.

This study is part of a user-centered design approach in the construction of an open archive platform, planned to create institutional repositories that will be managed by librarians in their respective institutions. Using this tool, the authors would be able to store, convert (XML), fully index, perpetuate, valorize, and distribute their digital documents.

For those using, managing and/or developing such platforms, the results of this study could be used in understanding lecturers and researchers behavior and expectations. Likewise, through the use of statistics derived from this study, it is a first step in weighing our convictions and formulating questions on future digital library users.

Introduction

A digital library platform, CASTOR (CApitalization & STORage) [1] has been developed since three years. Our work takes place in a user centered design approach to build an open-archive platform, planned to create institutional repositories, managed by librarians in their respective institutions. With this system, the authors are able to store, convert (XML), fully index, manage, perpetuate, valorize and distribute their digital documents. As we can't find any experimented users with this type of system in our institutions, the user centered design approach has focused on a study of users information environment. Even if we have a strong literacy on digital library use [2] [3], including local (French) studies [4][5], we need in a platform development cycle to capture the practices and the vision of future users, reflecting the local knowledge. The main objectives of the survey was to:

- Understand the documentary practice of lecturers and researchers (for a better usability)
- Take a census of users needs (for adapted services)
- Evaluate the future users ability to adopt the digital library system (to perform a better information and formation on this new service)

This survey has involved all the people researching and/or teaching at the Ecole Des Mines de Nantes: professors, associate professors, but also PhD students, research engineers, etc. These lecturers and researchers are divided into 5 departments: Automatic control and industrial engineering, Energetic and environmental engineering, Computer science, Subatomic physics and associated technologies, Human and social sciences.

Data were collected through an interview in order to agree with the interviewed person on the meaning of the concepts used. This was all the more useful as the interviewed sample badly knew the vocabulary used by librarians: the meaning of words, nowadays frequently used, remains unappreciated (for example: ICT), the meaning of a word can be different according to the interlocutor (for example: digital

library). The first part of the interview was a questionnaire [6] with open answers about the documentary practices. The second part was a semi-directing conversation dealing with the information expectations and needs of lecturers and researchers.

78 lecturers and researchers were met, they represent half of the Ecole Des Mines de Nantes lecturers and researchers. This sample consisted of volunteers who were free for the survey: no panel was defined as a preliminary.

The survey was designed and carried out in order to enable a double reading of the results: qualitative and quantitative. Data coming from the questionnaire were converted in percentages thanks to a simple statistical treatment, either directly, when the questions inferred short answers corresponding to the choice of an alternative, either indirectly, after a thematic regrouping of the ideas expressed. This last method was also used to analyze the conversations.

The lecturers and researchers documentary practices

Q .1 The beginning of an information retrieval

The information retrieval first aims at recovering full text documents, generally scientific articles: for 47 percents of the lecturers and researchers, searching for information means searching for documents. The remaining 53 percents expect either a document or a piece of information (for example a mathematical constant).

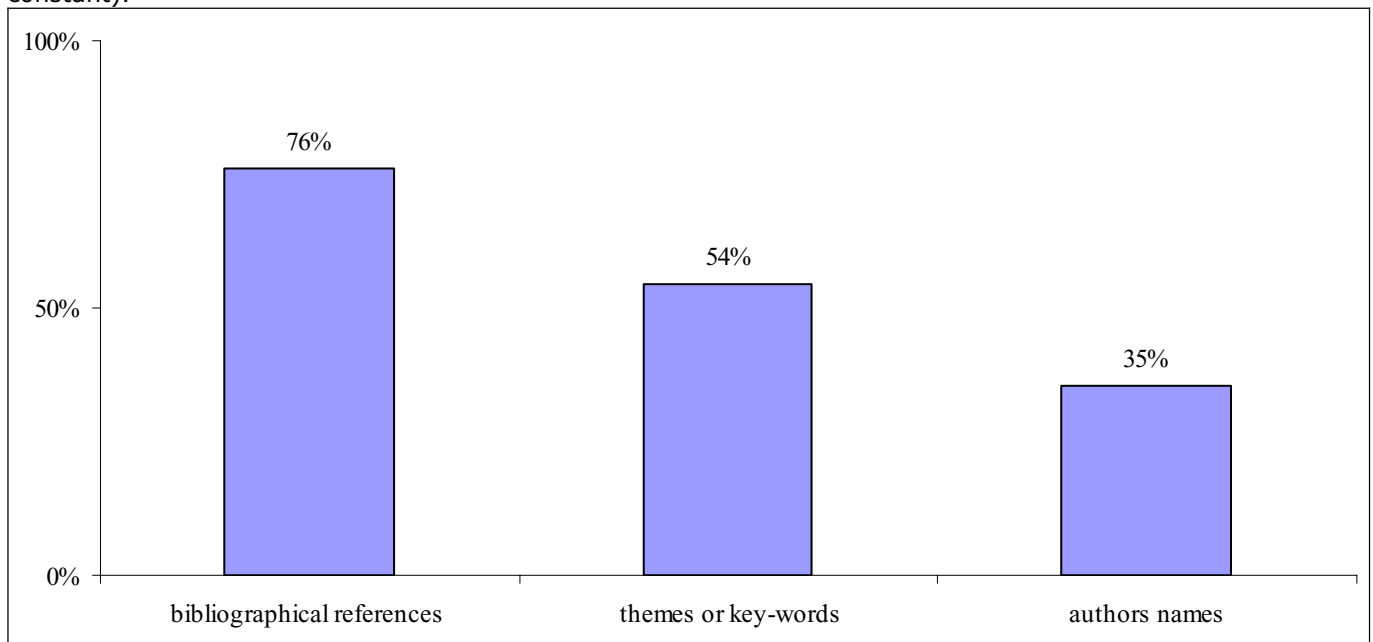


Figure 1: the starting points of information retrieval (% of lecturers and researchers starting to search using name(s) of author(s), theme(s) or keyword(s), bibliographical reference(s))

The three elements listed in figure 1 are often used successively: first, the lecturer and researcher searches widely starting from some topics or key-words ; then, when the subject is surrounded, he knows which authors write on the theme he's interested in and draws some new hints from the bibliographical references found in the selected documents.

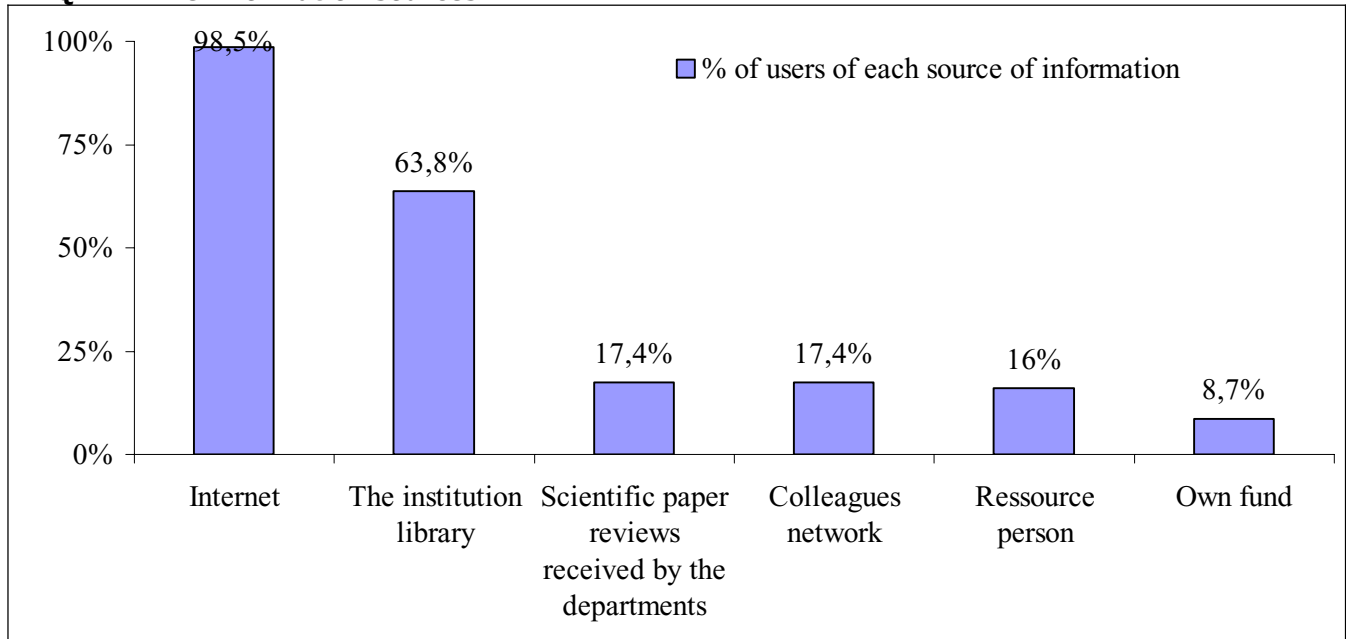
Q.2 The information sources

Figure 2: the main information sources

As the figure 2 shows it, every lecturer and researcher use Internet. More than 20 percents of them even only use Internet, unless exception. Others use more various sources.

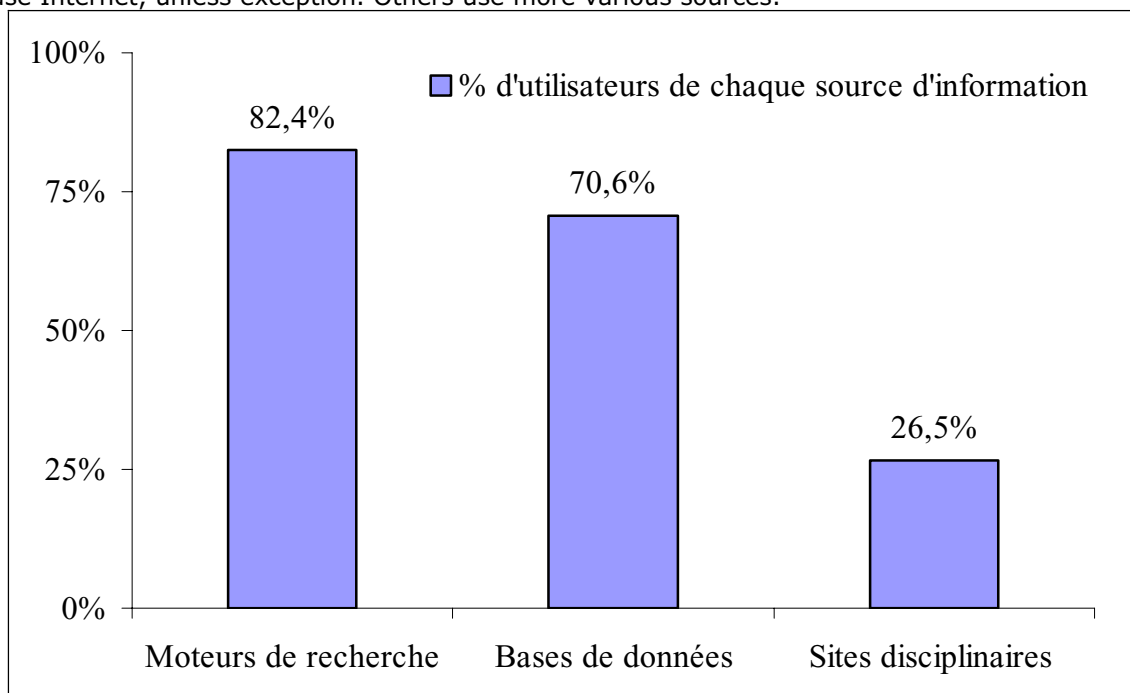


Figure 3: The Internet sources of information

The figure 3 gives the details of the main ways to reach the information available on Internet:

- the search engines, usually generic, Google ahead ;
- the data bases of bibliographical references and/or full text documents (with a preference for full text documents), appreciated for the ergonomics of their search interfaces ;
- the disciplinary websites (scientific review, university, institute, or research team websites etc.), especially for grey literature.

The paper resources are still important (figure 2): lecturers and researchers appreciate the paper resources in the Ecole Des Mines de Nantes library. This fund is supplemented by the departments' subscriptions. The colleagues network enables to get a document, whereas the resource person is used to have a track (for example: name of author, or review, website address, etc.) even bibliographical references. The lecturers and researchers personal documentary fund is not very much developed.

Several of them admitted that they prefer searching again on Internet a document that they already had printed than trying to find it in their office.

Q.3 The troubles with information retrieval

80 percents of the interviewees have difficulties in their research of information. Figure 4 gives the list and the distribution. These difficulties add to the inherent problems linked with information searching: it's not always easy for a person to define exactly what she wants, and to formulate the appropriate queries.

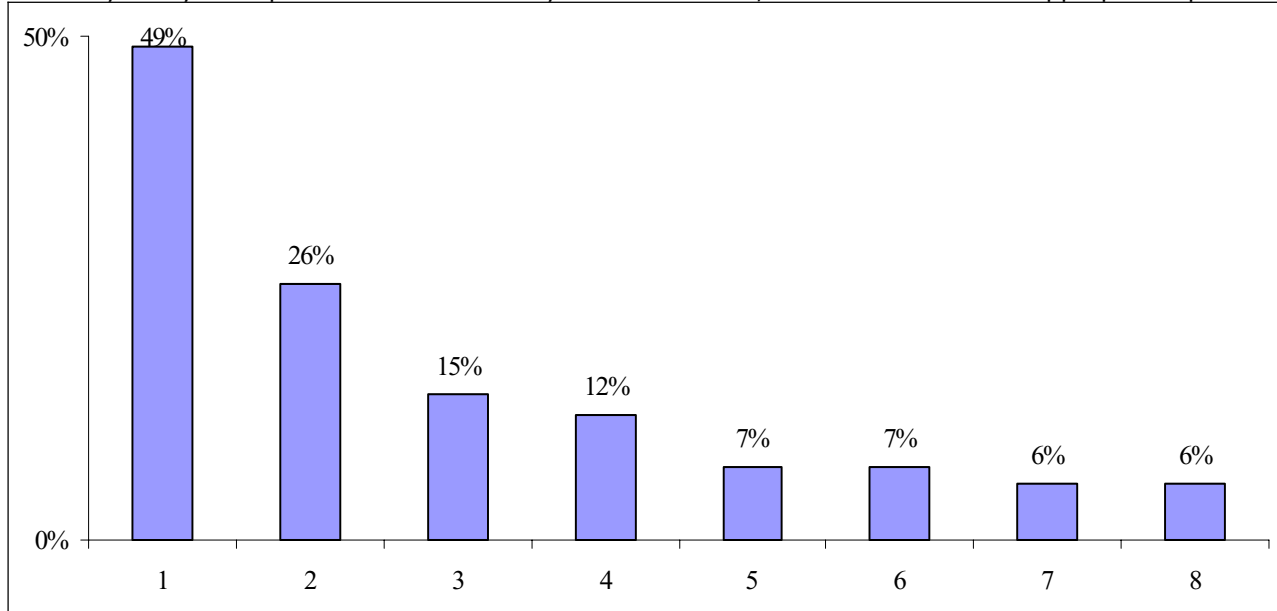


Figure 4: the difficulties of information searching (see legend below)

1. The lecturer and researcher doesn't find what he wants
2. There are too many answers in the middle of which relevant information is drowned
3. The lecturer and researcher lacks of information searching method
4. It takes a long time to search for information
5. The lecturer and researcher has to face a lot of technical problems (to reach Internet, to visualise documents, etc.)
6. Problems begin when the lecturer and researcher wants full text documents
7. It's difficult to find a few years old document
8. The lecturer and researcher must constantly adapt his query according to each website (its structure, the search engine operation, etc.)

The two main problems of the search for information, noise and silence, often correspond to two successive stages: first, he obtains too many answers, then, when he reduces its search, he may get only a few ones, even none. These results pose an interpretation problem: « *when there is no answer, how to be sure that there is actually no result ?* », « *how to be sure not to have missed any piece of information ?* », « *how to pick information without losing any other relevant one ?* ».

The interviewed lecturers and researchers often feel a lack of method while searching for information. Their knowledge is mostly empiric, and gives them the feeling of "managing" rather than really knowing how to search. Paradoxically, they admit not to try to ask for assistance.

The advent of Internet has induced some contrasted effects: on the one hand, we have a better access to information, on the other hand, there is a waste of time to select all the answers displayed by the generic search engines. Moreover, the increase in information supply also means an increasing number of the different interfaces that you have to get used of. This recent overflow of information hides the eternal problems of the access to full text or old documents (however relative oldness: a few years or decades). At last, Internet considerably makes the dependence to technical questions worse.

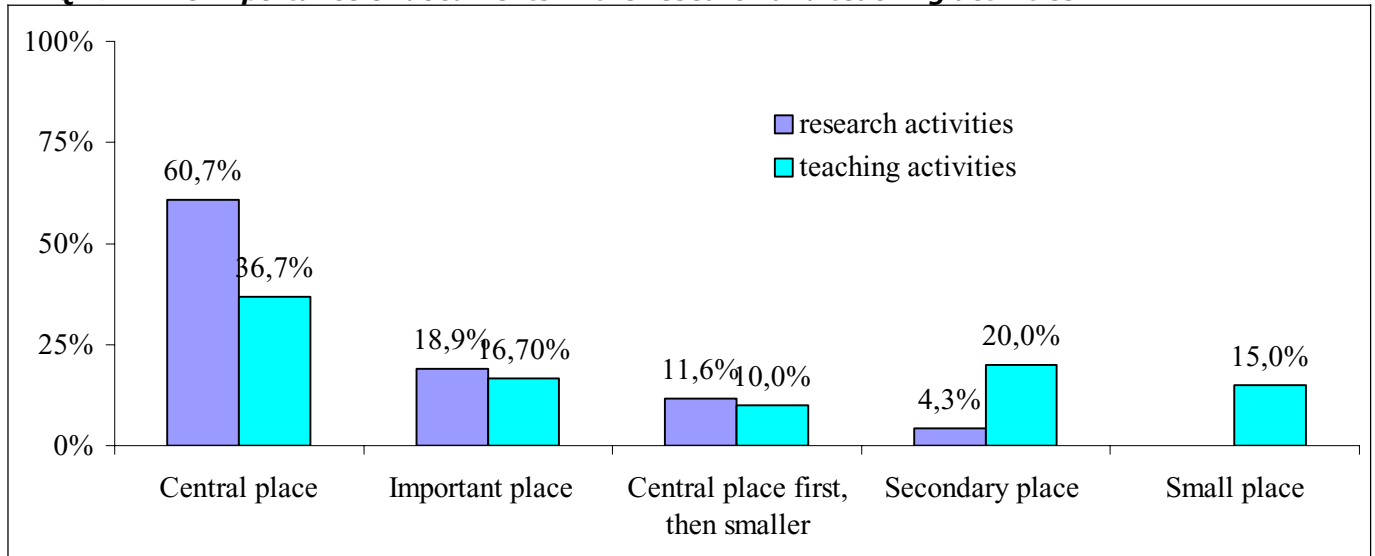
Q.4 The importance of documents in the research and teaching activities

Figure 5: the part of documents in the research and teaching activities

For teaching or research activities, documents, found after an information retrieval, play a main role according to the majority of lecturers and researchers. Many of them think they are central. Others « only » give them an important place: a document is only part of a research or teaching activity but is not central. Some lecturers and researchers think that the document's importance is first central then gets smaller. Indeed, at the beginning of a research, even at the time of a lecture creation, it is necessary to do the state of the art, then the personal contribution of the researcher become a priority and the documentary gathering evolves to watching and punctually searching for information.

Only a few lecturers and researchers give the document a small place. This group is especially composed of the tutors, often PhD students. In fact, in this case, the lecturer does the most part of the necessary information retrieval.

Globally, documents seem to be less important in the teaching activities than in the research activities. Several explanations can be proposed: the teaching subject is most of the time familiar to professors, and the transmitted knowledge evolves less quickly than in the research area.

Let's try now to appreciate specifically the place of the digital document (defined as a document displayed on a computer screen). The digital document use is general in research and almost general in the teaching activities

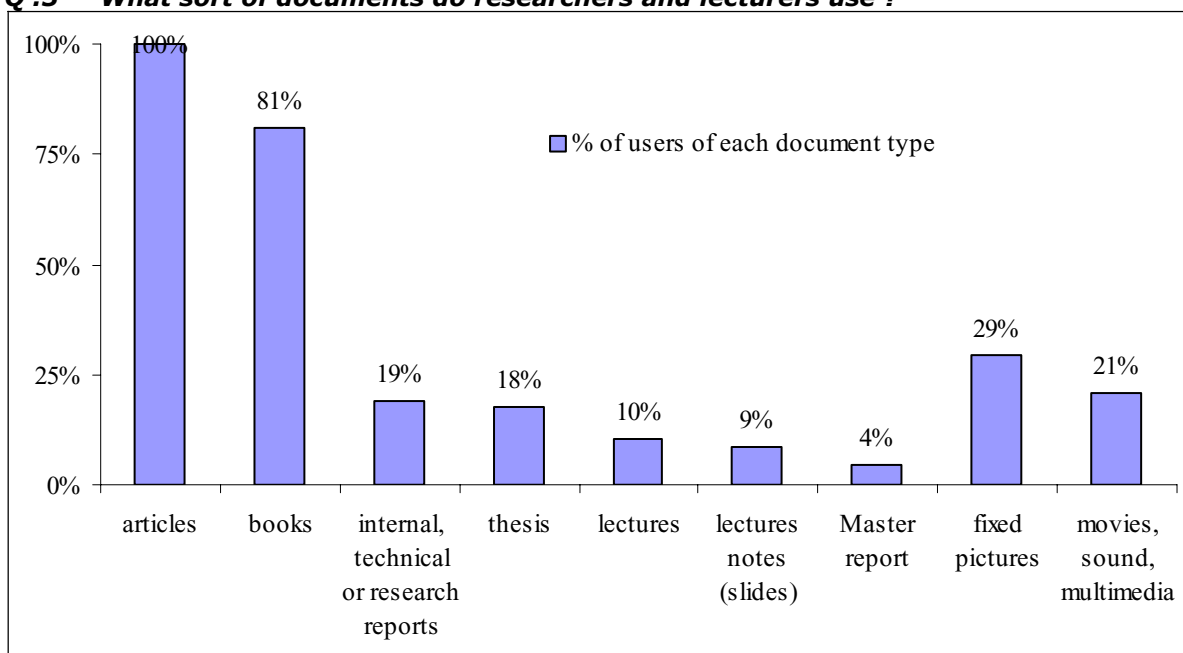
Q.5 What sort of documents do researchers and lecturers use ?

Figure 6: types of documents used, regardless of the medium

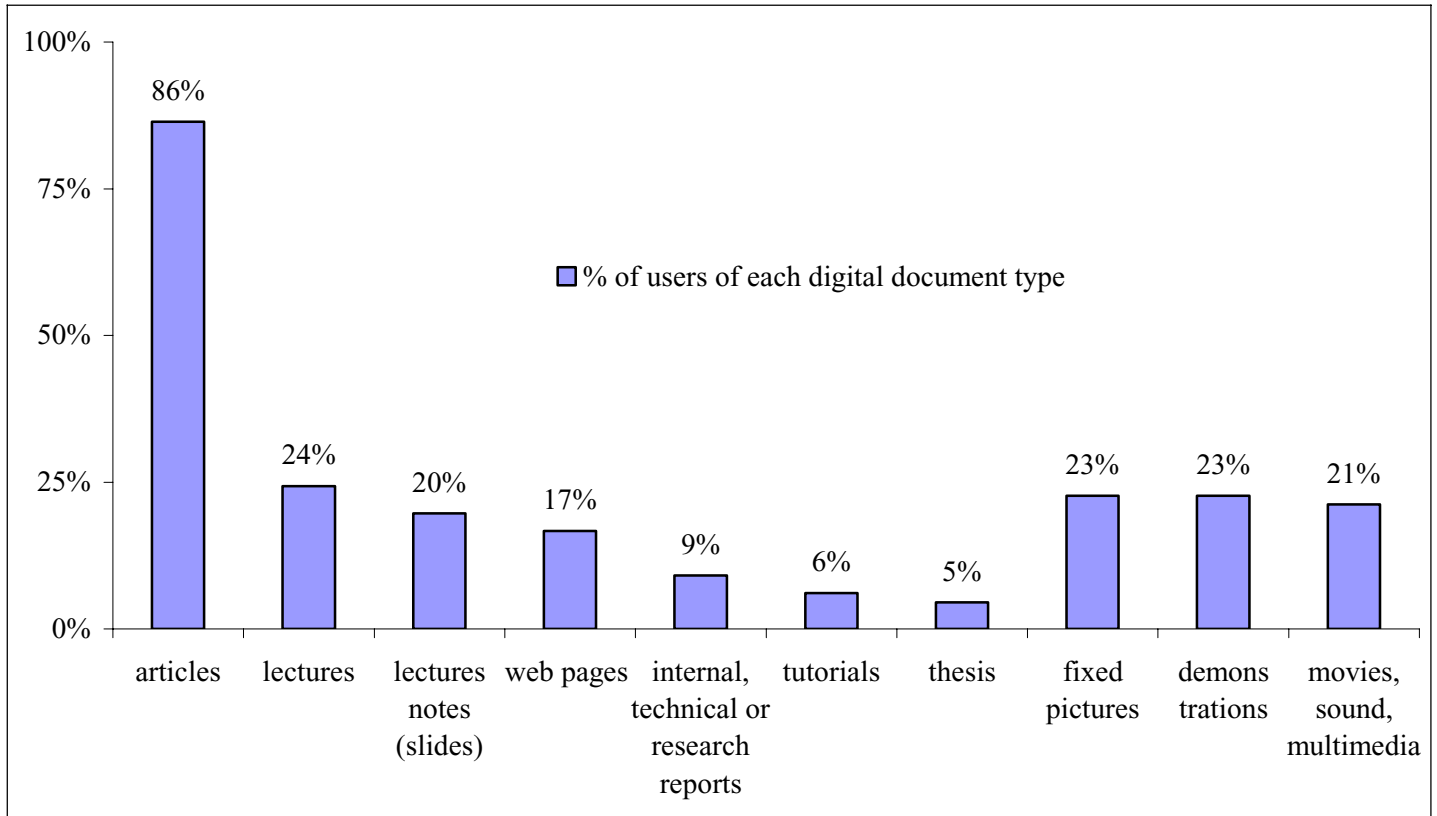


Figure 7: types of digital documents used

Whatever the document medium is, indifferent (figure 6), or specifically digital (figure 7), articles are by far the most used documents. As a general rule, these are scientific articles, sometimes, popularization work.

The comparison of figure 6 and 7 shows that digital reports and thesis are less used than reports and thesis in general, whatever the medium is. Digital books and Master reports are even missing. The relative small number of digital versions of these documents is a possible explanation. On the contrary, the more significant number of lectures and lectures notes in digital versions reflects the amount of the pedagogical resources available on Internet. At last, some specific digital types of documents appear: Web pages, tutorials, simulations and demonstrations (with short programs like java applets). The multimedia remains important: it is above all videos.

A sharper observation of the distribution of the different types of the documents used between research and teaching shows that:

- doing research, lecturers and researchers first use articles, then far behind books, at last thesis and reports
- while teaching they first use books, then far behind scientific or popularization articles, at last lectures and lectures notes

The same analysis restricted to digital documents shows that: articles form the overwhelming majority of the research resources ; but for teaching, lecturers and researchers use, in this order, lectures, lectures notes, articles and tutorials.

Q .6 What sort of documents do researchers and lecturers produce?

The production of digital documents is wide spread among the interviewed lecturers and researchers. However, these documents are still distributed by means of the paper. On the other hand, digital documents are not specifically designed for a digital distribution: scientific articles have the same form whatever the medium is. Indeed, when the document is distributed on a digital medium, as a general rule, this distribution is done by another person than the author: either because he hasn't his personal website, or because an editor distributes the document. The figure 8 details the types of digital documents produced and their distribution

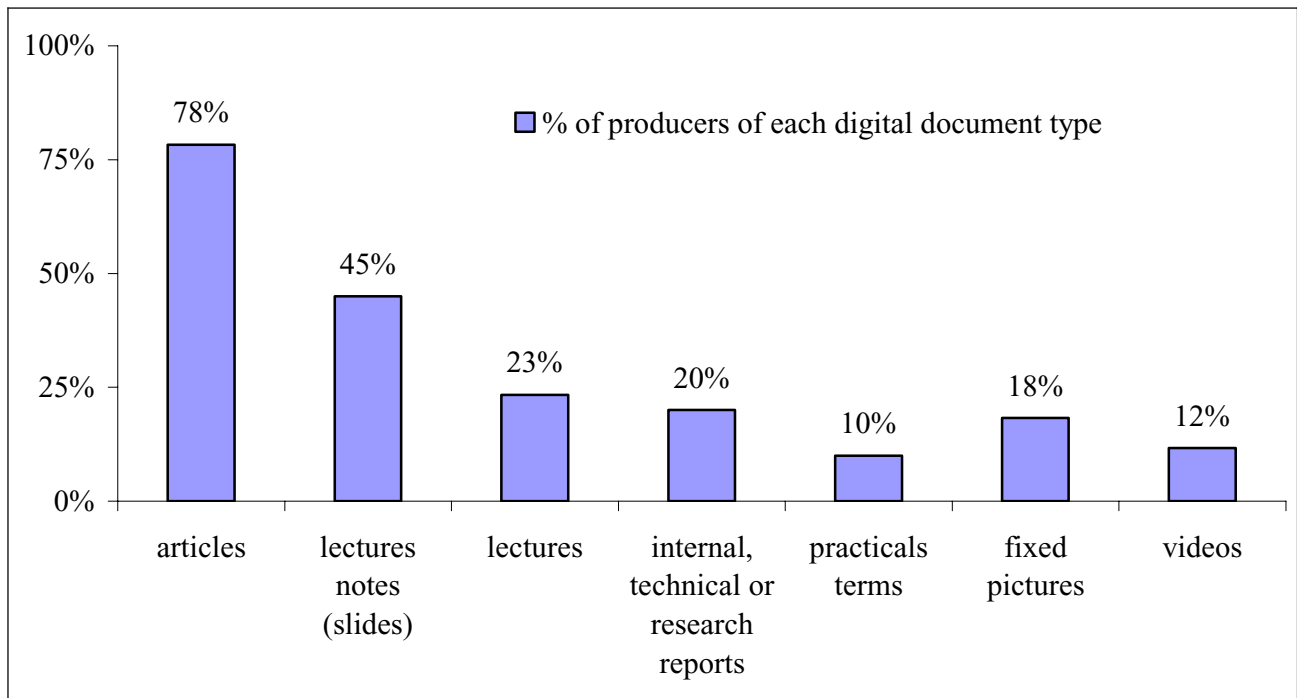


Figure 8: types of digital documents produced

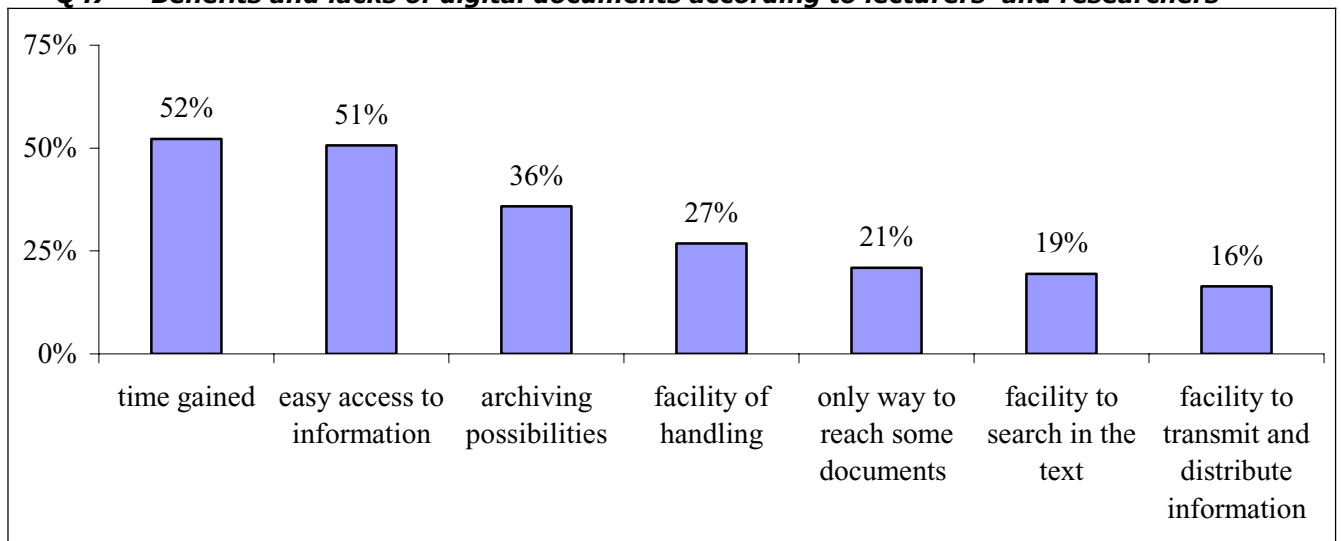
Q.7 Benefits and lacks of digital documents according to lecturers and researchers

Figure 9: the main benefits of digital documents

According to lecturers and researchers, the first benefit of a digital document is the time gained. Piece of information and documents are quickly accessible, almost without having to go out, and the lecturer and researcher is able to estimate the relevance of the obtained results faster than with paper documents. The second benefit is the easiness to reach digital information. Taking into account the difficulties encountered while searching for information, it may seem relative: according to lecturers and researchers, easiness means "to be able to reach a mass of very diverse information with some clicks". The archiving possibilities are the third benefit. In fact, a digitally stored document needs physically less space than the same document on paper. Moreover, it seems easier not only to organize a digital repository, using files, but also to search in such a repository. The facility of digital document handling is also appreciated: it's easy to get some extracts, have them modified, or extract some pictures, diagrams, etc, to insert them elsewhere.

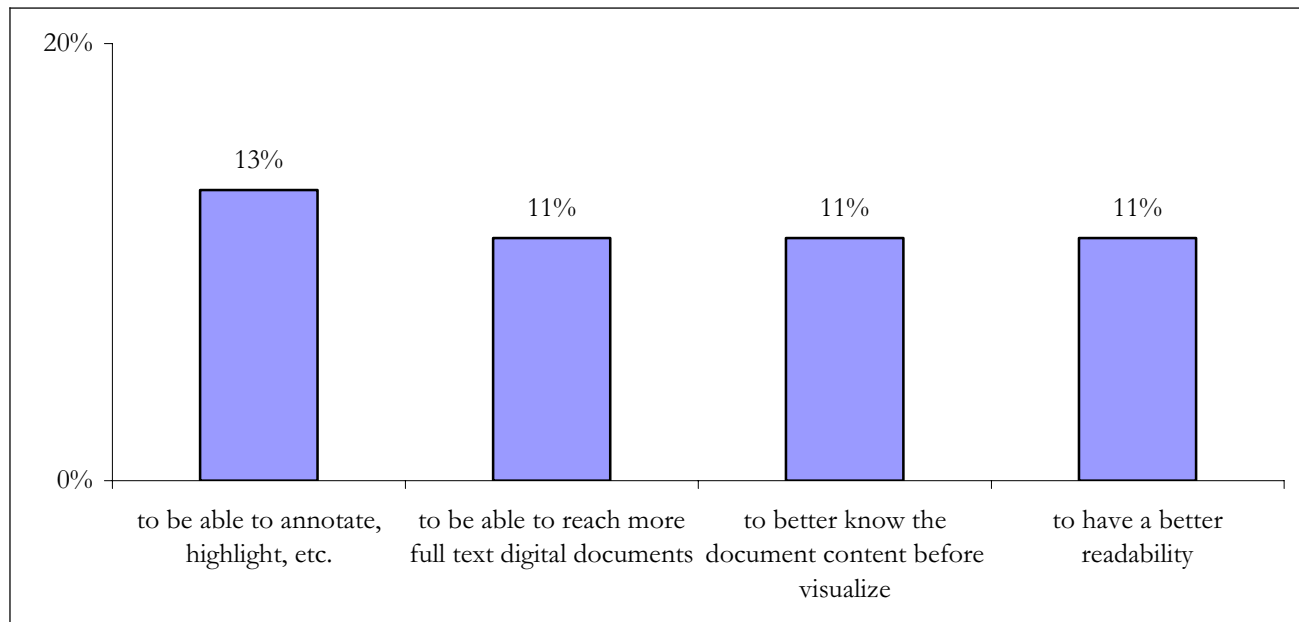


Figure 10: the mains lacks of digital documents

The first lack, according to the interviewed lecturers and researchers, is the impossibility to annotate, or highlight. But, this doesn't mean that they would want this possibility to be implemented: it only means that paper is irreplaceable. Lecturers and researchers would like to better know the document content before visualizing the full text: i.e. know its level (novice/advanced, popularization/research, etc.), have a summary, some key-words, and others meta-information (specifying the origin of the document, the film duration, the weight of the file to download, etc.). A lot of lecturers and researchers also wish a better readability: reading on a screen remains uncomfortable.

Synthesis

Nowadays, Internet has become a key-element in the lecturer and researcher information environment *"it's a window at the finger tip enabling to know what researchers do everywhere in the world"*. Digital technologies enable to find information quickly, to communicate, to send or to distribute documents easily, without moving from its office. So the lecturer and researcher have focused on the resources offered thanks to the computer.

In the same time, their perception of the role of the institution library seems to have become more vague, even if they still use the library services a lot. On the one hand, Internet creates a dazzling effect, which partly prevent from distinguishing the other information sources. On the other hand, lecturers and researchers have been upset when librarians, having to cope with Internet, sought to provide them autonomous ways of searching for information. Nowadays, it's not always clear for lecturers and researchers that some digital resources might only be available thanks to the institution library efforts.

Besides, it seems that there is a persistent problem of lack of training in information retrieval: a lot of lecturers and researchers have a limited practice of the Web potentialities.

Finally, lecturers and researchers seem to be somewhat over passed by the digital medium success: they find a lot of information on Internet, also receive a lot of information, but they have to face to relevance and reliability problems, and the time gained finding information is partly lost in selecting the results. But, although they are partly unsatisfied, the ratio between benefits and inconvenience remains interesting ; and lecturers and researchers are not yet really ready, or prepared, to spend time training to search for information.

Lecturers and researchers documentary expectations and needs

We present now the synthesis of the second part of the interview, the semi-directing conversation. Lecturers and researchers firstly wish to reach more digital resources. It means an easy and quick access to scientific information, full text documents, by means of targeted tools, which deliver reliable information and limit the time spent to select among the results.

Lecturers and researchers want first to be able to find easily the research documents, especially those produced in their institute *« because, even inside a laboratory, I don't necessarily know what my colleague does »*. It first means thesis, internal and research reports, and articles. Secondary, it means the documentary production of the students involved in a research team (research reports, training course reports), paying attention to quality and reliability. Lecturers and researchers also mention the pedagogical documents, especially those of their institute: lectures and other documents meant for students. Afterwards, they add the totality or a selection of the works made by students. Finally, bibliographies, experience reports, some numeric data (results or calculation checked files), data-processing code blocks, etc, could be capitalized.

Many of these resources are currently difficult to reach out of the direct neighboring of the author.

The interviewed public also thinks that it is important to rationalize the information access inside the institute: too much information risks to be missed because documents are scattered or lost. In the same time, it's necessary to ensure a long term access to resources.

A system with functionalities able to answer these expectations will however have to fulfill some conditions. An ease of use is a priority: « *even with less functionalities: the simpler the better* », « *the more knobs, the less visitors* ». The content must not only be rich and relevant but also structured in a clear and appropriate way. The interface must be user-friendly, quick and easy to reach. According to these points, lecturers and researchers think that an institute project of a digital library for grey literature is positive.

However, the survey conducted at the Ecole des Mines de Nantes also highlighted two potential obstacles to an optimum deployment of the project inside the institute. On the one hand, lecturers and researchers partly ignore their rights about the distribution of their documents. In the context of the development of the open access movement, it seems necessary to increase the authors awareness of the available possibilities. On the other hand, the lecturer and researcher's attitude may seem paradoxical. As a user, he wishes to get more full text resources. As an author, he may be reticent to allow a broad access to his documentary production.

The communication towards lecturers and researchers about this project will have to pay attention to these aspects and to prove that the digital library answers the expectations they had expressed. The platform specifications have also been built according to the survey results: it's the subject of the next part of this article.

Direct implications in digital library platform design

When the project has begun, we had three main hypothesis :

- First, we focus on the ability to manage any documents produced by students, lecturers or researchers. We did not limit only our development to popular added-value document such PhD Thesis.
- Second, librarians and their view of library have been put at the centre of the digital library design. The main implications result in advanced bibliographic description of documents, domain structured collections, controlled by librarians, and configurable workflow for documents submission that finish by a librarian validation (controlled keywords, collections, etc).
- Third, we used a user centered design approach (including this study, prototyping and evaluations) to design clear and simple user interfaces.

Beyond those directions, a lot of results from this study were used directly in the design of the digital library platform. Other results participate in refining the users picture we used to lead our choices. Some of them corresponded to well know behaviors of users, already depicted by our librarians in the early stages of the prototype design, and confirmed by real figures inferred by questionnaire results. But some results appear as new information on users practice and expectations. Following this study, in this last section, we are going to show some of the main changes we have made in our design.

Q .1 The beginning of an information retrieval

Doubtlessly, the question 1 confirms us that documents retrieval is a major preoccupation in users information seeking tasks .

Q .2 The information sources

Question 2 reveals that Internet and our institution library are the two main places for information retrieval (see Figure 2), but they appear far away from each other, strengthening the difference between paper and digital documents. Even if they manage digital library subscriptions, librarians are not seen as the reference people for digital information. In the idea of Grisemine [7], we build our digital library as a local service, an intermediate step between Internet and books-shelves. By means of such a tool, beyond the first goal of capitalizing document production, we hope to reduce the break between users *chaotic* internet practices and librarians role, bringing users to a more efficient and structured approach of document production and exploitation.

We can also see that lecturers and researchers use a broad variety of digital information sources, depending on a lot of parameters. A platform for aggregating the information of an institution like ours must find its place in their information space. So, we have developed a strong and clear position between existing (often commercial) digital library (ex: ScienceDirect, ACM DL, etc) and the diversity of document practices among users. The system results in two main elements:

- the submission / search document chain
- the profile management to store the users context (last seen documents, etc)

But those elements are relying on a software architecture based on open components, designed to be extensible and communicating with other services (LMS, other information systems, etc).

Q .3 The troubles with information retrieval

Question 3 helped us to choose between two opposed philosophies in document management system: the first one is only based on full text search, assuming that indexation engines have evolved a lot, and

on the other side, the metadata approach (still with full text search), that keep bibliographic records for each document, constraining users to fill notices. The second approach, more traditional, relies on controlled and free keywords, date of publication, etc, and allows to perform more accurate documents search, with better scores than full text search only (see Figure 4). In the future, users will be able to add semantic information to their documents, this should limit the amount of irrelevant information in their request.

Q .4 The importance of documents in the research and teaching activities

The question 4 shows us the importance of digital documents in these different activities (Figure 5). This question, among a few others, enabled us to draw the lecturers and researchers' profile, helping us to understand the evolution of their information environment. This profile has been useful all along the technical specification phase of the project, when we (often) had to make some little choices that changed the user interface or a task progress.

Q .5 What sort of documents do the researchers and lecturers use ?

According to question 5, lecturers and researchers use a wide variety of documents. They don't use digital books (Figure 7), and mainly manipulate lectures notes, research papers, and other grey documents. So, even if our first aim was to manage the largest variety of document types, this study revealed the real expectations, and we definitively focused on grey literacy. The main characteristics are a large production and many different types of documents based on many different types of stylesheets. Among them, a large part of 'hand made' digital documents doesn't fit any stylesheet, their structure rely only on the visual typography (character size, bold, etc). Users actually use document editors that allow them to structure their documents, but it is too early for semantic edition. It results in a global approach of document submission: while the author submits the document with an interactive approach, he fulfills a form with information about his document characteristics (stylesheet, format). It makes a strong difference for an XML conversion (compared to batch conversion). Given the heterogeneous mass of documents our users produce, we developed a generic approach to XML conversion, relying on document structure detection, even without any stylesheet. This conversion has been developed with Xerox XML experts.

Even if we try to have a very tolerant submission system, we need to teach users how to write a document, how to structure it and use a stylesheet. Only a few of our writers know a lot about this, especially new students.

Q .6 What sort of documents do the researchers and lecturers produce?

As seen before, lecturers and researchers use, but also produce a lot of documents (Figure 8). In question 6, they are as much actors as consumers of grey production, but we have to increase their awareness of document conservation, for research project continuation, for new comers, for external spreading, etc.

When an institute decides on grey literature conservation, their librarians get involved in both sides of information retrieval and production. It broadens their services and their importance. To be able to succeed in a digital library appropriation, librarians need a documentary policy, fixing the rules and encouraging the sense for the documentary patrimony of their institute.

Q .7 Benefits and lacks of digital documents according to searchers-lecturers

The last question confirms the interest of lecturers and researchers in document capitalization, and also gave us a few hints about useful services to propose or not (only 13,3% were interested in an *advanced* service like document annotating).

Conclusion

Finally, as we have seen above, this study had a great impact on the development of the platform. Since it has been released in November 2004, we are now starting evaluation tests with librarians and users to measure the performance and the qualities of the interface and the different services.

This survey, dealing with the Ecole des Mines de Nantes lecturers and researchers documentary practices, enabled to draw the portrait of a rather sedentary and solitary information seeker, focused on the resources delivered through his/her own computer. Documents, especially grey literature, play a main role in his/her research or teaching activities. Partly unsatisfied, this information seeker is aware of some problems: he/she is overwhelmed with information and lacks of method for information searching. However, the ratio between benefits and inconveniences of the current situation remains positive. Thus, lecturers and researchers are takers of intuitive tools but are not yet ready to spend time to improve their information retrieval method.

Acknowledgements

We would like to thanks the Pays de La Loire Council, and the French Ministry of Research for supporting this project. And a special thanks to Alice Ziernski for helping us to improve the legibility of this article.

Bibliography

- [1] the Castor project homepage, Ecole des Mines de Nantes. <http://www.emn.fr/bn/>
- [2] Gary Marchionini. *Information Seeking in Electronic Environments*, Cambridge Series on Human-Computer Interaction, Cambridge University Press, 2003.
- [3] Ann Peterson Bishop, Nancy A. Van House and Barbara P. Bittenfield. *Digital Library Use, Social Practice in Design and Evaluation*, MIT Press, 2003
- [4] CHARTRON Ghislaine. *Les chercheurs et la documentation numérique : nouveaux services et usages*, Paris: Editions du Cercle de la Librairie, 2002
- [5] SALAÜN Jean-Michel, VAN CUYCK Alain. *Les usages et les besoins des documents numériques dans l'enseignement supérieur et la recherche. Rapport final*, 1999, <http://www1.msh-paris.fr:8099/index.asp?choix=3-2>
- [6] the study questionnaire : <http://www.emn.fr/x-info/bn/documents/questionES.doc>
- [7] Marie-France CLAEREBOU, Jean-Bernard MARINO. *Grisemine, a digital library of grey university literature*, Fifth International Conference on Grey Literature (GL5), Amsterdam, 4-5 December 2003.

Author Information

Ms. Céline Bourasseau
EMN - Ecole des Mines de Nantes
4, avenue Alfred Kastler
44300 - Nantes
France

EML: celine.bourasseau@emn.fr
URL: <http://www.emn.fr/x-info/bn/>

Making Your Way Through Grey: Metadata, MARC and User Tools

Meagan Cooke, Senior Production Editor
Sean S. Costigan, Executive Editor¹

Abstract

Columbia International Affairs Online (CIAO) is the largest online collection of grey literature on international relations. Online since 1997, CIAO has grown to encompass both grey literature and published materials from over 200 contributing institutions. Currently the database contains over 10,000 papers and, as with anything of this magnitude, the challenge is not only continued growth in terms of content aggregation but also achieving successful user experiences and wider integration with other services. Our metadata is based on a small, controlled vocabulary that was developed in-house over the lifetime of the service. Pursuing an overhaul of CIAO's metadata for improved consistency has allowed for finer granularity in search results while also creating opportunities for the deployment of citation tools and rich cross-linking. In addition, the development of MARC records will allow for further integration with library resources and OPACs. This paper will discuss the impetus for the development of user and contextualization tools and our experiences in creating them.

CIAO's Background

CIAO is the largest library of international affairs content on the web. Originally funded by the Mellon Foundation, CIAO became self-sustaining through library subscriptions after three years of operation. CIAO was built in a partnership with Columbia's libraries, the University Press, and its academic computing and information systems group (AcIS). Subject specialists, computer scientists and librarians all had a hand in its initial development. Today, such expertise is drawn on to further realize the service's goals of promoting a wide range of grey and published literature in international affairs. Currently over 200 institutions partner with CIAO, primarily contributing working papers, conference proceedings, reports, books, policy briefs and journals. CIAO boasts more than 800 subscribers, among them government agencies, militaries, academic institutions and businesses. In any given month, over 2000 pages of material from dozens of contributors will be posted on CIAO. Such a large and mature repository poses significant challenges with regard to data management, archiving and customization.

Many Organizations, Many Standards

At CIAO's inception in 1997 a variety of file formats were commonly in use. CIAO's production staff was likely to receive files from Word, WordPerfect, Quark, and a smattering of non-standard text editors. In keeping with our desire to make CIAO usable to as extensive an audience as possible, all files were converted to faster loading html. Initially, CIAO adhered to HTML 2.0 specifications and when additional HTML specifications came out, adjustments were made to new, but not existing, content.

Today the bulk of CIAO's contributors deliver content in PDF (Portable Document Format) or Microsoft Word. For some of CIAO's subscribers, particularly those overseas and from secure locations, low bandwidth continues to be an issue that we design around by producing HTML abstracts for the majority of PDFs. In addition, html abstracts afford us the opportunity to more comprehensively describe the content using our metadata. Where possible we add author and title information to the PDFs, allowing our search engine to take advantage of that metadata as it indexes the site.

Metadata - A Primer

Several definitions of metadata exist. The W3C defines metadata as "machine understandable information for the web." For our purposes, metadata is simply "data about data" -- particularly information like keywords, document type, title, abstract, location, ISBN, etc.

¹ Meagan Cooke and Sean Costigan can be reached at mc1430@columbia and sc717@columbia.edu.

In its paper *Understanding Metadata*², the National Information Standards Organization details three main types of metadata:

- Descriptive metadata describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
- Structural metadata indicates how compound objects are put together, for example, how pages are ordered to form chapters.
- Administrative metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

To describe CIAO's content, our editorial team developed a metadata scheme that includes the following mixture of descriptive and administrative tags:

```
<meta name="robots" content="noarchive">
<meta name="ciao_title" content="">
<meta name="ciao_author" content="">
<meta name="ciao_type" content="">
<meta name="ciao_date" content="">
<meta name="ciao_subject" content="">
<meta name="ciao_subject" content="">
<meta name="ciao_subject" content="">
<meta name="ciao_subject" content="">
<meta name="ciao_institution" content="">
<meta name="ciao_language" content="">
```

Several renditions of these tags have been employed over the life of the service, due in part to technical and editorial developments that could not be described with the old criteria. Additionally, several hands have worked on the site since its founding, producing variations in standards and use. A production management system, built by CIAO's editors and web developers, has assisted in the standardization of metadata.

Controlled Vocabulary

Taxonomist Amy Warner defines a controlled vocabulary (CV) as "organized lists of words and phrases, or notation systems, that are used to initially tag content, and then to find it through navigation or search."³

Initially, nearly 100 subject tags (our controlled vocabulary) were chosen in consultation with CIAO's editorial advisory board. Over time these subject tags were distilled to around 80 in number and some new tags were added.

CIAO's subject tags are a mixture of regions, countries and themes. Four or more subject tags are chosen for each piece of content on CIAO, allowing users to find content in a variety of ways. Typically several region tags and thematic tags are used to describe documents. For CIAO, a complete tag might look like this:

```
<meta name="robots" content="noarchive">
<meta name="ciao_title" content="Final Report of the Independent Panel to Review DoD Detention
Operations">
<meta name="ciao_author" content="Schlesinger, James R.">
<meta name="ciao_author" content="Brown, Harold">
<meta name="ciao_author" content="Fowler, Tillie K.">
<meta name="ciao_author" content="Horner, Charles A.">
<meta name="ciao_author" content="Blackwell, James A. Jr.">
<meta name="ciao_type" content="wps">
<meta name="ciao_date" content="200408">
<meta name="ciao_subject" content="Crime">
<meta name="ciao_subject" content="International Law">
<meta name="ciao_subject" content="United States">
<meta name="ciao_subject" content="War">
<meta name="ciao_subject" content="Middle East">
<meta name="ciao_subject" content="Arab Countries">
<meta name="ciao_institution" content="U.S. Department of Defense">
```

² <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

³ <http://www.lexonomy.com/publications/aTaxonomyPrimer.html>

The Challenges of Inconsistencies

Insufficiencies have been addressed on two separate occasions through the introduction of new subject tags, though only after careful thought about the consequences. For a sobering analysis of metadata foibles, read Cory Doctorow's paper *Metacrap*.⁴ Since CIAO is built in a directory architecture with static html and PDFs, making new subject tags ripple through the site's directories can pose significant challenges. For example, how does one ensure that new tags are properly deployed in older content? With a database as large as CIAO, going through manually is clearly not an effective option, although files were modified manually over the course of the summer of 2002.

Due to the volume of content and the size of our staff, asking a university programmer, often from outside the immediate group, to write a script generally helps us come closer to our goals. After running such scripts, we proceed to manual verification and modification of stray oddities. Initially all work is done on a test server, with the eyes of the editorial and production staff (as well as committed librarians) watching for inconsistencies. After achieving general agreement that all is well, we find a time to copy the changed files over to the production server. Since CIAO has an international subscriber base, we pay careful attention to usage patterns around the world in order to minimize downtime.

Over the life of the service, we've made two additions to CIAO's administrative metadata. The first dealt with one of the ramifications of allowing Google to index the site. Shortly after allowing Google in, we realized that its robots had cached the site's pages. Adding a new administrative tag `<meta name="robots" content="noarchive">` and requesting that Google crawl the site again restored our security. Additionally, taking advantage of Google's prominence has greatly increased the number of free trials worldwide and brought users from subscribing institutions to CIAO. Usage has doubled.

In consultation with our editorial advisors, CIAO is now striking agreements with international institutions to collect non-English language content. With an eye towards searching non-English content we have added a tag noting its language. The default language remains English, though we are aggregating content in French, Spanish, German, Romanian, Turkish and are considering additional languages. All foreign language content includes abstracts in English.

Cataloging CIAO Content

In some disciplines the knowledge or information published in grey literature will never appear in published format⁵ (Chilag 1982). The Internet has gone a long way towards making the reports, studies and proceedings that constitute grey literature easy to access. According to Luzi⁶ (2000), grey literature databases were distributed commercially as early as the 1970s. To advertise these highly-valued, vetted resources we've worked with librarians and programmers in creating MARC records and disseminating citations. MARC is the acronym for MACHine-Readable Cataloging, an initiative that began over thirty years ago. Developed under the auspices of the Library of Congress, MARC formats are standards for "the representation and communication of bibliographic and related information in machine-readable form."⁷

When libraries subscribe to an electronic resource, the institutional libraries staff often has time to create one MARC record for the electronic resource, but not the materials it holds. Increasingly, academic libraries expect vendors to provide MARC records for materials within the resource. The benefit to libraries is that users can find publications across many resources through one search in the library OPAC (Online Public Access Catalog). This method is preferred to using popular search engines, such as Google, which index materials not vetted by information professionals. Users looking for a report will learn that the full text of that report is available electronically through their library's subscription to CIAO. Publishers benefit from the addition of many more access points to their resources and the likelihood that the files within those resources will be used increases.

In February 2003 we began working with the special materials cataloging staff at Columbia University libraries to catalog the published and grey literature in CIAO. As of November 2004, CIAO holds more than thirty full-text journals and 124 full text books. Because of the relatively manageable quantities of published literature, we decided that books and journals could be cataloged by library school students. Among the grey materials on CIAO, case studies were the only collection that could be cataloged manually. Library students cataloged CIAO case studies as part of their school practicum; the hours spent cataloging fulfilled their requirement for graduation.

⁴ <http://www.well.com/~doctorow/metacrap.htm>

⁵ Chilag, J. *Non-conventional literature in agriculture – an Overview*. 1982. IAALD Quarterly Bulletin, Vol. 27, No 1.

⁶ Luzi, Daniela. *Trends and Evolution in the Development of Grey Literature: A Review*. 2000. The International Journal on Grey Literature, Vol 1, Issue 3. <http://www.emeraldinsight.com/rpsv/cgi-bin/linker?reqidx=cw/mcb/14666189/v1n3/s2/p106.idx&lkey=-1038598568&rkey=911673>

⁷ <http://www.loc.gov/marc/>

The remaining grey literature collections in CIAO are vast; as of November 2004, the database holds more than 4,500 working papers and more than 5,000 policy briefs. Catalogers mapped descriptive metadata in CIAO abstract pages to MARC fields. Subjects in CIAO's pages were mapped to Library of Congress subject headings. With these values mapped to MARC fields and LoC standards, programmers are working to crosswalk CIAO metadata to XML and MARC.

We explored different models for distributing MARC records. CIAO case studies were uploaded to OCLC (Online Computer Library Center) and subscribing libraries were made aware of their availability through e-mail, allowing them to find these records through a title search in OCLC. As a distribution model, this was unsatisfactory for CIAO users and for CIAO staff. Catalogers at subscribing institutions found it time-consuming to search for the records and download them. Also, we did not have the ability to learn about this service's usage. As a result we chose instead to disseminate records from CIAO servers.

The MARC records for CIAO journals are now available and records for CIAO books will become available this spring. We expect the vast majority of these institutions to have OPACs, so usage in the majority of subscribing institutions should increase once MARC records for CIAO content are introduced. Sixty-five percent of CIAO subscribers are higher-education institutions. Instead of one access point to CIAO—the sole MARC record for the electronic resources—users will find more than 10,000.

Metadata Standards

Several standardization initiatives exist. For our purposes we have considered converting to Dublin Core. According to the Dublin Core Metadata Initiative's website, Dublin Core is "an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models." Dublin Core was intended to be a simple set of elements that could be used to describe web resources. Currently 15 elements make up the standard: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights. At this time CIAO continues to explore the value of converting its metadata to Dublin Core. Mapping our metadata remains the most likely outcome. Such mapping is likely to take place should we harvest the metadata for an OAI server.⁸

Rich Hyperlinking

It is certain that students comprise our largest user group. Specifically with students in mind, CIAO's editorial staff recently completed an atlas comprised of country data, maps and histories from the CIA, Department of State, Transitions Online and other sources. Database driven, the new atlas allows for data comparisons across a dozen variables. As an information resource, the atlas is rich enough to stand on its own, although we believe its true value will be realized when maps and country data are made available from within CIAO's content. The materials in the Atlas provide context for the technical grey literature on the site. To that end, we are currently introducing links to regional and country maps to all content in the working papers and books sections of the site. A link in the upperleft corner of a paper or book about a country or region will open a map of the region. As with other changes on the site, scripts have been written to locate subject tags within the content and introduce the appropriate map links. We chose to add this functionality to working papers and books first because more often these materials given unique subject tags whereas materials appearing in serial form—including policy briefs and journals—are given tags according to the journal or policy brief subjects, not the article content. Without clean and standardized metadata, this important development would not have been possible.

E-mail this Citation

Once the metadata on CIAO was made consistent through editorial labor and programmer scripts, we began to create functionality based on the information. Of course, we looked forward to more and better search results but we also wanted users to be able to cite the content they found.

CIAO includes a page devoted to guidelines for citing content found on the website. These guidelines were taken from the Columbia Guide to Online Style (1998). We went a step further, embedding an "e-mail this citation" link in all reports, conference proceedings and working papers on CIAO. The "email this citation" script examines the metadata from a given page and extracts the relevant information. It then puts this information together into a citation. On the user end, clicking on the "email this citation" link

⁸ "The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content." OAI FAQ (www.openarchives.org)

renders a form in a new window. Users who complete the form are able to send citations for CIAO content in emails to a specified address. Users can email citations that adhere to APA and Chicago styles. It is hoped creation of this tool will facilitate usage of papers and reports; and advertise their availability to researchers and scholars.

Search

Columbia University employs an Inktomi search software product called Ultraseek. Ultraseek was purchased by AcIS in June 1999 as a replacement for other search engines in use at the university. Prior to 2001, CIAO used a homebrew search engine that was fast but had one glaring weakness: it was unable to index PDFs. When more contributors began delivering content in PDF it became clear that our technology was no longer sustainable. CIAO's staff began searching for a replacement search engine and we realized that, as a member of the university community, CIAO could benefit from Columbia's site license. Since its initial deployment on CIAO we have completed three modifications of the search form, each time learning a little bit more about Ultraseek's capabilities and shortcomings.

Ultraseek is programmed in Python, a little known language at Columbia, and so changes are made infrequently by AcIS programmers. Over the years we have made modifications to the search form based on requests from librarians and metadata cleanup initiatives. We have experimented with adjusting the weight of different search criteria, including title, body and subject metadata. Currently, criteria for searching include publication format, author, contributing institution, region, title, subject and date. At this time we cannot sort publications by date, but we are investigating other search technologies, including Lucene and Google, that can perform that function. The CIAO language metatag will be used when we have a larger body of foreign language content.

The Future is Grey

In the next year, we will migrate to an XML-based site architecture. The promise of XML is that it will allow content to be easily modified while also separating content from its metadata scheme. Implementing site wide changes will be less labor intensive. We will also be able to break individual documents into their components, offering greater granularity in search results. The ability to search abstracts is an often-requested feature that will be included in the future XML version of CIAO. Prior to migrating to XML we are likely to harvest all CIAO's metadata and build it into an OAI server.⁹ Such initiatives promise to further our goals of disseminating international affairs grey literature.

Further Reading

1. *Understanding Metadata* <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
2. *Grey Literature: Its History, Definition, Acquisition and Cataloguing*
<http://www.moyak.com/researcher/resume/papers/var7mkmkw.html>
3. *Grey Literature and Library and Information Studies: A Global Perspective*
<http://www.emeraldinsight.com/rpsv/cgibin/linker?reqidx=/cw/mcb/14666189/v1n4/s3/p167.idx&lkey=-1810531140&rkey=785358>
4. *The Expanding Horizon of Grey Literature*
<http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/glpaper.htm>
5. *Metadata Resources* <http://www.ukoln.ac.uk/metadata/resources>
6. *Dublin Core Metadata Initiative*, www.dublincore.org
7. OAI, www.openarchives.org

Author Information

Ms. Meagan Cooke,
Production Editor
Columbia University Press
61 W. 62nd Street
New York, NY 10023
USA

Email : mc1430@columbia.edu
URLs : www.ciaonet.org
www.earthscape.org

⁹ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Relating Intellectual Property Products to the Corporate Context

Keith G Jeffery CCLRC, Anne Asserson UiB

Abstract

The knowledge society demands that organizations utilise to the full the IP (intellectual property) they generate and acquire. In a context of research, development and innovation, the IP consists of products, patents and publications (in the widest sense – any stored representation of human intellect). While conventional research publications (white literature) provide much of the visible IP, the ‘submerged part of the iceberg’ is the grey literature. With OA (Open Access) technologies the line between white and grey becomes blurred, because of accessible postprints and preprints. Significantly, white literature (and grey literature) may commonly be hyperlinked to further grey literature in the form of technical reports, procedure manuals, learning materials, software, data (in all media), communications (e.g. e- and voice-mail) etc. The hypothesis is in three parts:

- a) hypermedia: hyperlinks from white (and grey) literature to the ‘submerged iceberg’ of grey publications enriches and completes the IP of the original publication;
- (b) CRIS: (Current Research Information Systems) that cross-link the material to other entities in the business domain (where the business here considered is R&D) provides a context to enable the end-user to utilize more effectively the publication or collection of publications for their purpose;
- (c) Metadata: of high quality for controlling integrity, providing access, providing a description and managing rights is essential.

At both UiB, Norway and CCLRC, UK systems have been built utilising CERIF (Common European Research Information Format which provides a data structure to match (b) above) linked with publications information. Both are based on joint research by the authors, the initial design results having been reported in GL’99 (Jeffery 1999). Both implementations have demonstrated effectiveness with enthusiastic end-users. The UiB system is targeted at providing UiB with research productivity information and uses a relational database structure extending CERIF 2000 with bibliographic entities/attributes. This allows publications to be related to people, to organizational units and to projects. At CCLRC, using relational database technology, a CERIF-based extended corporate data repository (including people, organizational units and projects) is linked with an OA eprints institutional archive, containing postprints and preprints (and also other grey publications). In each case, the relationships are annotated with role (such as author or reviewer) and start date/time, end date/time. The relationships can be many-to-many, thus a publication may have many authors each of which may have affiliations with many organizational units and projects. Each instance of organizational unit, project or person may be recursively related to another; this allows correct representation of hierarchic or network (connected graph) relationships. For both implementations the next steps involve implementing fully the formalised Dublin Core metadata proposed initially by the authors in GL’99 and since developed further by experience (Asserson and Jeffery 2004). This will provide easier syntactic (structural) interoperability and allow semantic interoperability. However, more importantly this development will allow the end-user to utilise the virtual collection, whatever the media and whatever the quality, in a knowledgeable and context-aware way.

1 BACKGROUND

In a context of research, development and innovation, the IP (Intellectual Property) consists of products, patents and publications (in the widest sense – any stored representation of human intellect). While conventional research publications (white literature) provide much of the visible IP, the ‘submerged part of the iceberg’ is the organisation’s grey literature. This commonly represents its ‘know how’ or knowledge base (Jeffery 1999), (Jeffery, Asserson et al 2000). There are also legal considerations: many organizations protect their IPR with patents or pre-publication; copyright and database right are counterbalanced by Freedom of Information and Data Protection legislation. Innovation, Technology Transfer, Wealth Creation, Quality of Life are major objectives of R&D, and the reason why national governments, commercial organizations, charitable organizations and even individuals invest in it. Most of the technology upon which we depend today is the result of R&D years ago, and similarly the quality of life we enjoy is largely the result of R&D in topics such as medicine, education, environment. This is the IP.

The WWW (World Wide Web) (W3C), has made e-publishing inexpensive and easy. This has led to an explosive growth of institutional repositories. The Open Access Initiative (OAI) utilised the Dublin Core (DC) metadata standard and harvesting software (OAI-PMH) to link the repositories. The two great challenges for the Web as outlined in (Berners-Lee 1999) are the semantic web (to make the web understandable) and the web of trust (to make it secure). The semantic web is now being constructed, largely by (a) more formal data structures which are suitable for manipulation by first order logic, commonly involving the use of structured metadata and (b) use of domain ontologies to provide

definitions of the meanings of terms and the logical inter-relationship of terms – as supportive associative metadata. The web of trust implies that material is secure from misuse and that the organization holding the information is trusted to utilise it in a way coincident with business ethics. This is achieved by associative restrictive metadata related to the original information.

Metadata is data about data; a classification was proposed (Jeffery, 2000) which ensures separation of metadata kinds and assists in correct logical processing. The application of the metadata classification to CRISs was demonstrated in (Jeffery, Lopatenko et al 2002).

The GRIDs concept – initially metacomputing i.e. linking supercomputers (Foster and Kesselman 1998) - has been extended (initial internal papers in 1999 and published in Jeffery 2001) to a full-blown distributed computing environment including, as GRID services, the W3C concept of web services together with concepts of the semantic web and web of trust. This environment provides the platform for the ultimate blurring of white and grey literature, from refereed publications through annotated preprints to technical reports and manuals all cross-linked to datasets and – using CRIS technology (www.eurocris.org) - to persons, organizations, projects, patents, publications, events, facilities and equipment. The environment is completed with associated computation power, special output facilities (e.g. VR (Virtual Reality) and dynamic control of detectors and instrumentation collecting data.

The key to an IP-management environment is based on three related concepts:

(a) hypermedia: comfortable hyperlinks from white (and grey) literature to the 'submerged iceberg' of grey publications enriches and completes the IP of the original publication. Hypertext systems evolved through the last 40 or 50 years with visionaries such as Ted Nelson leading the way (Xanadu). The extension to through multimedia to hypermedia is demonstrated in Microcosm (MI) and Hyperwave (HW);

(b) CRIS: cross-linking the material to other entities in the business domain (where the business here considered is R&D) provides a context to enable the end-user to utilise more effectively the publication or collection of publications for their purpose. The key results are reported in (Jeffery 1999), (Asserson Jeffery et al 2002), (Asserson and Jeffery 2004) ;

(c) Metadata: of high quality for controlling integrity, providing access, providing a description and managing rights is essential. The importance of metadata in a modern GRIDs environment was described in (Jeffery 2000) and in a CRIS environment in (Jeffery 2004a).

These three characteristics – if implemented correctly - can provide an environment to relate grey literature to the corporate context and so improve the corporate knowledge value of the grey literature. Here lies the major thesis of this paper.

2 IP IN A CORPORATE CONTEXT

Each organization needs to utilise its IP for business benefit (including the business of R&D) and for public relations / marketing purposes. This implies that an organisation needs to know the IP it owns, catalogue it, curate it and understand the business benefits from it – not least to inform investment decisions in future R&D to generate further IP. The knowledge base consists not only of the white literature but also the 'iceberg' of grey literature encapsulating the know-how of the organization in technical reports, instruction manuals, training materials etc. Furthermore, increasingly the IP rests in datasets (e.g. results of drug clinical tests), in databases (e.g. customer relationship information) and in software (which encapsulates the business processes of the organization). The whole may be subject to scrutiny through audit or freedom of information requests and thus there are great incentives for an organization to manage well these IP assets.

In the R&D world most public funding bodies now assess the output – IP - from an organisation that it has funded - e.g. a university or research institute - and they base future funding decisions – at least partly – on that information. Thus recording, curation and management of IP is critically important to research-based organisations. Similarly, for many modern businesses the quality of its IP determines success and future investment from shareholders. Thus we derive the requirement: to provide systems and an environment such that organisations can manage effectively their IP bringing together both the IP itself (e.g. grey literature) and the organisational business structures and objectives.

3 SOLUTION

3.1 Components

The three components required to satisfy the requirement have been identified above: hypermedia, CRIS and metadata. Here we equate components to existing systems and future work.

3.1.1 Hypermedia: Open Access Repositories

The WWW (World Wide Web) (W3C), has made e-publishing inexpensive and easy. Subject-based repositories started to come into existence (e.g. ArXiv(ArXiv)) sustained by the community of researchers who deposited copies of pre- and post-prints in such repositories, in the latter case in addition to publication in a journal (paper or e-). Individual authors posted pre- and post-print publications on their own web pages, challenging publisher copyright and making material available freely. Certain teams produced software to crawl the web and produce catalogs of publications from personal pages.

Organisations then created institutional repositories where employees could deposit their publications and where the organisation could claim the IPR and gain the publicity. The Open Access Initiative (OAI) utilised the Dublin Core (DC) metadata standard and harvesting software (OAI-PMH) to link the repositories. A recent UK Parliamentary enquiry supports institutional repositories for publicly funded research.

3.1.2 CRIS: CERIF

CERIF (www.eurocris.org/cerif) has been developed over a number of years by a nominated team of national experts operating in a EU (European Union) context. The key developments are summarized in (Asserson, Jeffery et al 2002). Briefly, CERIF provides a formal datamodel for CRIS which is flexible, extensible and interoperable. Furthermore, the datamodel also provides the generic basis for a corporate data repository as demonstrated by CCLRC.

3.1.3 Metadata

Metadata is pervasive through both the above components. The thesis is that formal metadata, processable by logic, can improve greatly the coupling of grey literature, via CRIS to the organisational objectives for IP management.

3.2 The Marriage

3.2.1 UiB Solution

UiB (University of Bergen) is using a system for cataloguing, as metadata, Grey Literature and other publications. The experience qualifies Anne Asserson as a member of the board developing a national system. The comparison of the UiB datamodel with the datamodel proposed in (Jeffery 1999) was described in (Jeffery, Asserson et al 2000).

3.2.2 CCLRC Solution

CCLRC is using CERIF as the basis for a corporate data repository to manage not only IP assets but also to control access to corporate data resources, to control business processes and to provide management information for decision support. The aim – in addition to making CCLRC business operations (including scientific R&D) as effective and efficient as possible - includes being able to evaluate R&D outputs in terms of persons, organizations, corporate governance, cost, time and quality. This is achieved by crosslinking an OAI-compliant Institutional Repository with the CERIF-compliant corporate data repository.

3.2.3 The Proposed Solution

Since CERIF2000 in 1999 – which only provided a 'hook' to link a publication (stored in another database) to CERIF - the authors have been working to produce a CERIF-compatible extension to handle publications in such a way that: a) the end-user has convenient access to the metadata and ideally onward to the publication itself wherever possible; b) the end-user can treat statistically the metadata representing the publications for management purposes; c) the field of conventional publishing is represented but also grey literature; d) hyperlinked multimedia data is handled.

3.2.3.1 The Original CERIF-Compatible Proposal for Bibliographic Extension

This is perhaps best represented by a table comparing 'normal DC' with the 'formalised DC'

DC	Formalised DC
	<UNIQUEID> RAL92-003 </UNIQUEID>
<TITLE> A Distributed Architecture to Provide Uniform Access to Pre-Existing Independent, Heterogeneous Information Systems </TITLE>	<TITLE> <language> en </language> <title> A Distributed Architecture to Provide Uniform Access to Pre-Existing Independent, Heterogeneous Information Systems </title> </TITLE>
<CREATOR> Naldi F, Jeffery K G, Bordogna G, Lay J O, Vannini-Parenti I </CREATOR>	<PERSON><role>author</role><person>Naldi F</person></PERSON> <PERSON><role>author</role><person>Jeffery K G</person></PERSON> <PERSON><role>author</role><person>Bordogna G</person></PERSON> <PERSON><role>author</role><person>Lay J O</person></PERSON> <PERSON><role>author</role><person>Vannini-Parenti I </person> </PERSON>
<SUBJECT>Current Research Information Systems; legacy; heterogeneous; distributed; protocol; communications; data; exchange</SUBJECT>	<SUBJECT><language>en</language> <scheme> RALClassification </scheme> <subject> Current Research Information Systems </subject> </SUBJECT>
	<KEYWORDS> <language> en </language> <scheme> UKThesaurus </scheme> <keywords> legacy; heterogeneous; distributed; protocol; communications; data;

	exchange </keywords> </KEYWORDS>
<DESCRIPTION>A system named EXIRPTS has been built which demonstrates access over distributed multilingual information systems of R&D projects. The system resolves problems of resource location and utilises a catalog technique for metadata which allows the end-user to have a homogenous view over heterogeneous information</DESCRIPTION>	<DESCRIPTION> <language> en </language> <description> A system named EXIRPTS has been built which demonstrates access over distributed multilingual information systems of R&D projects. The system resolves problems of resource location and utilises a catalog technique for metadata which allows the end-user to have a homogenous view over heterogeneous information </description> </DESCRIPTION>
<PUBLISHER>Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX UK </PUBLISHER>	<ORGUNIT><role>publisher</role><orgunit>Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX UK </orgunit> </ORGUNIT>
< CONTRIBUTOR> Wright, L, Daniels,T </CONTRIBUTOR>	<PERSON> <role> contributor </role> <person> Wright, L </person> <role> proofreader </role> <person> Daniels, T </person> </PERSON>
<DATE>1992</DATE>	<COVERAGE TEMPORAL> <project> 1988-1991 </project> <publication> 1992 </publication> </COVERAGE TEMPORAL>
<TYPE>Technical Report</TYPE>	<RESOURCE TYPE> <scheme> RALLibrary </scheme> <language> en </language> <resourcetype> Technical Report </resourcetype> </RESOURCE TYPE>
<FORMAT>Word2</FORMAT>	(note handled by conventional MIME typing)
<IDENTIFIER>RAL 92-003</IDENTIFIER>	<RESOURCE IDENTIFIER> <scheme> RALLibrary </scheme> <resourceidentifier> RAL92-003 </resourceidentifier> <scheme> referencelist </scheme> <resourceidentifier> [NaJeBoLaVa92] </resourceidentifier> </RESOURCE IDENTIFIER>
<SOURCE > [null]	Note: done using relationships between resources referenced by UniqueId
<RELATION> [JeLaMiZaNiVa89] </RELATION>	<uniqueid> <RAL92-003> </uniqueid> <role> preliminary investigation </role> <uniqueid> [JeLaMiZaNiVa89] </uniqueid>
<COVERAGE> Europe,1983-1991 </COVERAGE>	<COVERAGE SPATIAL> <scheme> LatLong </scheme> <coordinates>10W35N-30E80N </coordinates> <precision> 5degrees </precision> </COVERAGE SPATIAL> <COVERAGE TEMPORAL> <scheme> years </scheme> <constraints> [1983<x>1991] </constraints> </COVERAGE TEMPORAL>
<RIGHTS> Copyright Rutherford Appleton Laboratory 1992 </RIGHTS>	(note handled separately with access, privacy security etc)

This table demonstrates clearly the additional formalisation with language and schemes used to make precise each element where applicable. Furthermore repeating groups of values are handled consistently and the more complex relationships such as publication <-> person (in roles author, editor, contributor) are handled correctly and flexibly using a consistent method (binary linking relations) as in the rest of CERIF.

3.2.3.2 The New Bibliographic Extension

Since this work there have been developments, to make more precise the associative restrictive metadata and the associative descriptive metadata for classification and evaluation in the light of further experience of actual and potential uses:

a) more work has been done on the restrictive associative metadata and the following is proposed:

DC	Formalised DC
	<UNIQUEID> RAL92-003 </ UNIQUEID >
<RIGHTS>	<UNIQUEID> <SECURITYSCHEME> <SECURITYLEVELCONSTRAINTS>
<RIGHTS>	< UNIQUEID > <PRIVACYScheme> <PRIVACYLEVELCONSTRAINTS>
<RIGHTS>	< UNIQUEID > <ACCESSRIGHTSScheme> <ACCESSLEVELCONSTRAINTS>
<RIGHTS>	< UNIQUEID > <CHARGINGScheme> <CHARGEAMOUNT>

WHERE <CHARGEAMOUNT> is extended with sub-elements as follows:

<CHARGEAMOUNT> <currency> <amount> </CHARGEAMOUNT>

b) more work has been done on quality assessment metadata and in particular an annotation capability, especially for eprints, and the following is proposed:

DC	Formalised DC
	<UNIQUEID> RAL92-003 </ UNIQUEID >
	< UNIQUEID > <relation> <ANNOTATION> <PERSON>

where <ANNOTATION> is extended with sub-elements as follows:

<ANNOTATION> <language> <representation> <format> </ANNOTATION>

and <PERSON> may be extended with the additional subfield <DSig> for digital signature.

c) additional work has been done on quality assessment metadata to allow the end-user to do analyses of publication performance. The following is proposed:

DC	Formalised DC
	<UNIQUEID> RAL92-003 </ UNIQUEID >
	< UNIQUEID > <CLASSIFICATIONScheme> <CLASSIFICATIONVALUE>

Where <Classification Scheme> may be ISI SCI (scientific citation index) or weblinks to the page or anything else. This is integrated with the general classification scheme over all CERIF attributes. These new entities have as their primary key <UNIQUEID> which actually identifies the metadata record, NOT the publication itself being described by the metadata, which is a foreign key <RESOURCEIDENTIFIER> in the formalized DC. Clearly the structure (Fig 1) allows the designer to add as much or as little metadata content as required depending upon the application.

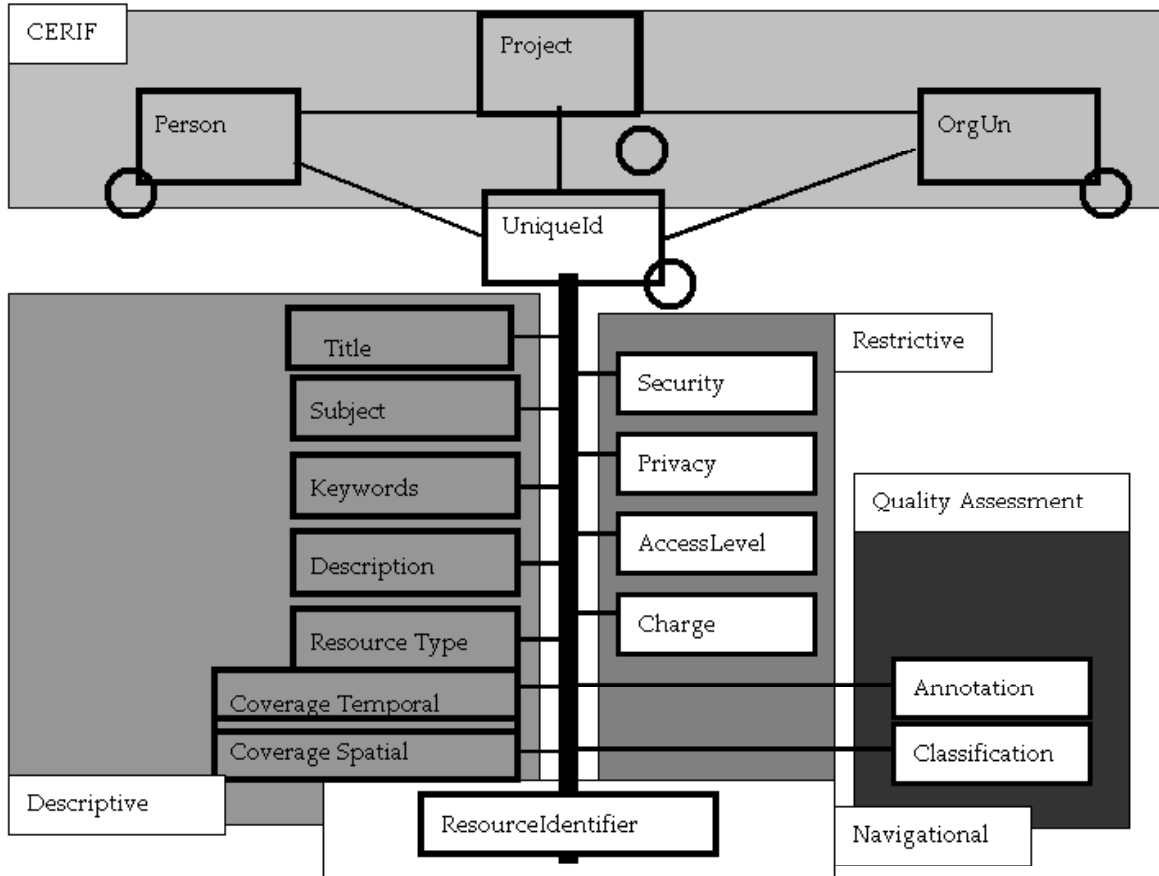


Fig 1: Proposed CERIF Publication Metadata Record based on Formalised Dublin Core

4 THE WAY FORWARD

This scheme is now being implemented progressively at CCLRC to assess its utility in practice. UiB had already an implemented system extending CERIF to record publications specifically for their purposes of publication assessment. However, UiB is assessing the data model proposed here with a view to migrating to it. Meantime, the DC community has extended progressively 'vanilla' DC with 'qualifiers'. However, these provide essentially attributes and thesaurus terms to be used for semantic interoperability and do not address the syntactic (structural) issues, nor the semantics of links between data entities.

Author Information

Prof. Keith G. Jeffery,
Director Information Technology
Department for Computation and Information
CCLRC - Rutherford Appleton Laboratory
Chilton Didcot, Oxfordshire OX11 0QX
United Kingdom

EML: kgj@rl.ac.uk
EML: k.g.jeffery@rl.ac.uk
URL: <http://www.itd.clrc.ac.uk>
URL: <http://www.itd.clrc.ac.uk/Person/K.G.Jeffery>
CRIS2004 <http://www.eurocris.org/conferences/cris2004/>
euroCRIS <http://www.eurocris.org/>

5 REFERENCES

(ArXiv) www.arxiv.org

(Asserson Jeffrey et al 2002) Asserson,A; Jeffery,K.G; Lopatenko,A: 'CERIF: Past, Present and Future' in Adamczak,W & Nase,A (Eds): Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 33-40 2002 (available under www.eurocris.org)

(Asserson and Jeffery 2004) Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 29-40 (available under www.eurocris.org)

(Berners-Lee 1999) Berners-Lee,T; 'Weaving the Web' 256 pp Harper, San Francisco September 1999 ISBN 0062515861

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

(Foster and Kesselman 1998) Foster I and Kesselman C (Eds). The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kaufman 1998

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

(Jeffery 1999) Jeffery, K G: 'An Architecture for Grey Literature in a R&D Context' Proceedings GL'99 (Grey Literature) Conference Washington DC October 1999 <http://www.konbib.nl/grey-net/frame4.htm>

(Jeffery, Asserson et al 2000) Jeffery, K G; Asserson, A; Revheim, J; Konupek, H: 'CRIS, Grey Literature and the Knowledge Society' Proceedings CRIS2000 Conference, Helsinki, Finland May 2000 <http://www.cordis.lu/cris/cris2000>

(Jeffery, 2000). Jeffery, K G: 'Metadata': in Brinkkemper,J; Lindencrona,E; Solvberg,A (Eds): 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0.

(Jeffery 2001) Jeffery,K.G.; 'GRIDs: Next Generation Technologies for the Internet' Invited Keynote Presentation (Abstract in Proceedings) Eds Wang,Y; Patel,S; Johnston,R.H; OOIS2001 Conference, Calgary August 2001, page 1, Springer, ISBN 1-85233-546-7

(Jeffery, Lopatenko et al 2002) Jeffery,K.G; Lopatenko,A; Asserson,A: 'Comparative Study of Metadata for Scientific Information: The Place of CERIF in CRISs and Scientific Repositories' in Adamczak,W & Nase,A (Eds): Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 77-86 (available under www.eurocris.org)

(Jeffery 2004) Jeffery, K.G.; 'GRIDs, Databases and Information Systems Engineering Research' in Bertino,E; Christodoulakis,S; Plexousakis,D; Christophies,V; Koubarakis,M; Bohm,K; Ferrari,E (Eds) Advances in Database Technology - EDBT 2004 Springer LNCS2992 pp3-16 ISBN 3-540-21200-0 March 2004

(Jeffery 2004a) Jeffery, K.G.; 'The New Technologies: can CRISs Benefit' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 77-88 (available under www.eurocris.org)

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

(OAI) www.openarchives.org

(W3C) www.w3.org

(Xanadu) <http://jefferson.village.virginia.edu/elab/hfl0155.html>

Chasing the Grey Evidence: A standardised Systematic Critical Literature Review approach

P SIMKHADA¹, J KNIGHT,¹ G REID, L WYNESS, L MANDAVA,

(¹ Presenting authors and authors for correspondence)



Partners for Research in
Emergency Nutrition



UNIVERSITY OF ABERDEEN

Department of Public Health
University of Aberdeen

Abstract

Goal: This paper advocates that the triangulation of evidence to give a direction for planning should include the distillation of experience and information that abounds in unpublished reports. Evaluating complex interventions using only high level of evidence from peer reviewed published papers may limit the ability to assess effectiveness, however there is a lack of a standard robust approach to reviewing the grey literature. The way forward is the development of a 'Systematic Critical Literature Review' SCLR (which was originally developed for trial data), applicable to such grey literature, which systematically collates the literature and has a validated framework to assess the quality of the evidence.

Methodology: To assess the need for developing a review methodology for grey literature, - a Grey SCLR, a standard SCLR on published literature on childhood malnutrition in humanitarian complex emergencies was undertaken and assessed for completion of the evidence base. A mechanism of systematically collating the grey literature within this subject was developed and yield assessed. This method was based on standard SCLR methodology and values, and key terms from a previous standard SCLR on published literature carried out by our research group were used. Secondary references and expert opinions were also sought. The theoretical quality assurance literature was reviewed to form a basis towards developing a quality framework applicable across the different types of un-published reports and studies.

Results/Conclusions: A standard SCLR is theoretically applicable to complex situations. But in such situations an evidence base cannot be created by even a highly sensitive standard SCLR on published literature. The majority of information is in the grey literature and a review of such should have the main values and the key steps equivalent to a standard SCLR for published literature.

Three stages were developed: 1. Review of peer-reviewed publications; 2. Application of key words to web based grey literature search; 3. Further refinement of search following identification of key organisations through stage 2; 4. Search through experts within key organisations identified. The value of the evidence found was assessed against the stages of the methodology developed and against the key steps of a standard SCLR.

The theories of quality assurance gave key concepts, which were developed into a quality framework for the un-published literature. This was then piloted and reliability and context validity sought. Grey literature is an essential part of the evidence base for practice in complex interventions, which may have multi-stakeholders, have multi-variables, have a lack of predictability and robust data and require a broad literature approach. An outcome from a standard SCLR in such a complex field is not robust and so this 'Grey SCLR' was further developed to increase the required effectiveness and breadth of a literature review, but based on the objectivity of a standard SCLR. A developed Grey SCLR should be transferable to inform evaluations of other complex interventions and can complement a review of the published studies.

Introduction:

Evidence based practice in public health medicine is now developing its own methodology.^{1,2,3,4} Evidence based public health is the informed, explicit and judicious use of evidence that has been derived from any of a variety of science and social science research and evaluation methods.⁵ Humanitarian non-governmental organisations increasingly acknowledge and require such evidence. Humanitarian complex emergencies show starkly the challenge of obtaining the appropriate evidence for evidence-based practice.

The standard rigorous systematic review for peer reviewed publication, with its values of being comprehensive, transparent, reproducible and having minimum bias, is the benchmark to collate and analysis current evidence. But in complex public health situations, effective, appropriate research for policy decisions will necessarily involve research designs other than randomised control trials. The above definition of evidence based public health highlights two main issues particularly relevant for complex situations and complex interventions: publication bias and insufficient evidence on process and context for policy decisions and implementation.

There is now a move from strict hierarchical evidence to a focus on evidence aligned with quality and appropriateness, therefore multiple methodological studies are required to be included in a review of the literature for an evaluation of complex programmes.^{2,3, 4,6,7,8} But multiple methodologies, lower levels of the hierarchy of study designs, monitoring surveys, studies from developing countries, studies in a field with low priority re publication, classic historical work and new studies (publication time lags) may all contribute to publication bias and so incomplete evidence.

Scientific evidence cannot stand on its own, quality assessed evidence on the geopolitical, social, cultural contexts are required and the acknowledgement of previous perceptions at all levels of the programme, for sufficient evidence to inform policy decisions and enable effective implementation.^{2,3,6,7}

In the field of humanitarian nutrition programmes in complex emergencies there is therefore a need to focus on the 'grey' (literature that is not the 'white' literature on peer-reviewed literature databases) in order to actually minimise publication bias and to include valuable process evaluation and the ability to place the findings in context. What is required is a comprehensive systematic search and quality assessment of the grey literature, these requirements are applicable to most complex situations.^{2,4,9,10} The grey literature in the humanitarian nutrition field will include reports and studies from all organisations involved including global, national and local levels and key experts, theses, conference abstracts, working papers etc.

The advantages of including the grey literature is that it gives breadth to the printed evidence reducing publication bias through access to new research and a wider range of research designs and also with the world wide web. It contains process information for policy makers and local information for planning and implementation. The disadvantages are that it requires rigorous quality control - as it may lack a formal peer-review and also strict inclusion criteria, with the large volume of grey literature available. But both these aspects are also required for a standard systematic review of peer-reviewed literature.

An agreed systematic search methodology is thus required for reviewing the grey literature that incorporates the values of a gold standard review of published literature without its problems of publication bias and producing insufficient evidence for policy decisions and implementation.

A pyramid or wedge identifying representative organisation, key experts, databases and web sites, at different levels from global policy to field work, has been formed through findings from the review of peer reviewed literature and web search. The consensus validity of the representativeness can be validated through key experts at all levels and a systematic, reproducible, explicit and comprehensive (within defined inclusion criteria) search undertaken.

Pilots for a search methodology were undertaken which highlighted the challenges for the next stage of the development of the methodology. Complex situations by their nature have multiple stakeholders with different priorities; the workforce in complex emergencies is very mobile making contact difficult; the relevant web sites gave many challenges to systematically filter valid and reliable grey literature; grey databases were found to be not widely subscribed to and their completeness were questioned.

Frameworks for the quality assessment for including and integrating best available literature, which will include lower hierarchy designs and qualitative and descriptive studies are prevalent in recent literature, two recent examples are the TREND and GRADE statements.^{7,11} But there is still the need for a consensus on appropriateness and relevance to the available grey literature in complex emergencies and in addition any quality assessment requires an agreed evaluation theoretical basis relevant to this field.¹²

Partnership working between academia and humanitarian non-governmental organisations ensures the appropriate skill mix of robust academic techniques will take account of relevant and sufficient

process and context information required for policy and implementation. This partnership is now taking forward the problems found through the initial stages of developing an appropriate framework for a systematic grey literature review methodology, which can mirror the rigorousness and validity of a standard systematic review for peer – reviewed published literature and be based on an agreed evaluation theory. The identification of such valuable grey literature can be time consuming and difficult, such a framework aims to make the identification feasible for the evidence required for policy and implementation decisions required in other complex situations.

References:

- 1 Walters E. and Doyle J. Evidence based public health practice: improving the quality and quantity of the evidence. *Journal of Public Health Medicine* 2002; Vol. 24:3
- 2 Walters, J., Doyle, J. and Jackson, J. Evidence based public health: improving the relevance of Cochrane Collaboration systematic reviews to global public health priorities. *Journal of Public Health Medicine* 2003 vol 25 (3) 263-266.
- 3 Victora, C. *et al.* Evidence based public health moving beyond randomised trials. *American Journal of Public Health* 2004; Vol.94:3
- 4 Kelly *et al.* *Getting evidence into public practice*. Health Development Agency 2004; www.hda.nhs.uk
- 5 Sacket, D. *et al.* Evidence based medicine: what it is and what it isn't. *British Medical Journal* 1996; 312: 71-72
- 6 Howes, F. *et al.* Evidence based public health: the importance of finding 'difficult to locate' public health and health promotion intervention studies for systematic reviews. *Journal of Public Health* 2004; Vol.26:1
- 7 Des Jarlais, DC, *et al.* Improving the reported quality of non-randomised evaluations of behavioural and public health interventions. *American Journal Public Health* 2004; Vol. 94,3
- 8 Hawe P. *et al.* Complex interventions: how 'out of control' can a Randomised Control trial be? *British Medical Journal* 2004; Vol.328
- 9 Oliver, S. *et al.* An emerging framework for integrating different types of evidence in systematic reviews for public policy. *Paper presented at methodologies sub group of the HAD: Public Health Evidence Steering Group* 3 February 2003
10. Kaptchuk, TJ. Effect of interpretive bias on research evidence. *British Medical Journal* 2003; 326:1453-1455
11. GRADE working group. Grading quality of evidence and strength of recommendations. *British Medical Journal* 2004; Vol. 328:1490-1494
12. Pawson R. and Tilley N. *Realistic Evaluation* 1997; SAGE Publications Ltd., London

Author Information

Dr. Jane Knight
Univeristy of Aberdeen
Public Health Medical Advisor Food Standards Agency Scotland
Public Health, Polwarth Building, Foresterhill
Aberdeen AB25 2ZD
United Kingdom

email: j.knight@abdn.ac.uk
http://www.abdn.ac.uk/public_health

The integration of GL documents with the research information system on occupational safety and health

**Daniela Luzi, CNR/IRPPS, Maria Castriotta, ISPESL
Mariarosaria Manco and Renza Mazzucco, LINK s.r.l., Roma**

Abstract

The paper presents the results of a joint project between the Italian National Institute of Safety and Health at Work and the Italian National Research Council. It illustrates the choices and main features in the development of the RIS-OSH system (Research Information System on Occupational Safety and Health). This system has been developed giving particular attention to the organisational improvement of the production, collection, preservation and diffusion of institutionally generated information, emphasising the importance of the quality control of the information produced during the project lifecycle. The paper gives the description of the workflow, which outlines the interaction between the various actors involved in the process of supplying and diffusing results. Finally the integration of the RIS-OSH system with the GL database and the global system architecture are described.

Introduction

In recent years a growing number of Italian scientific institutions - universities, research organisations - and governmental bodies have started using network technologies, in order to make information concerning their projects available with modes that range from the availability of web pages to the consultation of databases, as well as the retrieval of the documents resulting from the research projects. These systems have various purposes; they satisfy the need to diffuse results within the scientific community, become useful instruments for technological transfer to potential users of the results and, in a broader sense, to society, which plays an increasingly active role in scientific policy making. As the assessment issues take hold in Italy, these systems also assume ever-greater importance since they make it possible to perform qualitative and quantitative analyses that are of use in making choices concerning research policy.

In 1997, the Italian National Institute of Safety and Health at Work (ISPESL) set up a database containing approximately 500 records referring to projects carried out between 1996 and 1998. This archive was based on a "static concept", i.e. the data was entered when projects had already reached an advanced phase and updates were made difficult by the lack of a constant, co-ordinated data flow. On the basis of this past experience, and of the critical points that emerged from the update procedures adopted, a totally new product was planned to enable the exploitation of new technologies and to redesign the organisational model; thus improving the information flow connected to the creation of a research database. For this reason, in 2003 a joint ISPESL - CNR (the Italian National Research Council) project was launched, which had two main objectives:

- to analyse the process of activation of ISPESL research projects and identify the technologies best suited to improving it, with particular regard to work organisation and result diffusion, as well as to develop an information system on research projects funded by ISPESL to external commissioning parties;
- to improve the diffusion of information on research through the integration of information on projects with the other electronic archives that constitute ISPESL's documentation patrimony. A first step towards this integration was made by connecting the new system with the pre-existing database on Grey Literature (GL), (<http://www.ispesl.it/lg/default.htm>).

The first objective aims at attaining organisational and technical-scientific advantages. On the one hand improvements in inter and intra-departmental communication are achieved; consistency and standardisation of documentation relating to the project are assured; administration procedures are simplified by reductions in the time required to activate projects and diffuse information, thus also improving process control and monitoring. On the other, the advantages expected relate to an enhancement in the Institute's visibility and prestige; the possibility of performing scientific and economic assessment of research activities and the availability of an instrument for future research activity programming.

The second objective, that of integrating the research information system (hereinafter referred to as RIS-OSH, Research Information System in Occupational Safety and Health) with the database on GL, stems from the need to make the best possible use of the products of research. They are primarily constituted by the deliverables presented at the end of the projects, i.e. those GL documents that are identified as Activity Reports in the GL database. The explicit relationship between the description of the project and its results has the primary advantage of improving the complex process of technological transfer, favoured by the diffusion of the information contained in the deliverables. The

integration between the two environments also facilitates the retrieval of information, thus allowing a more direct updating of the electronic GL database.

The paper presents the results of the joint project and illustrates the choices and main features of the system developed, which was based on the broader context of the research information systems on which it was modelled. The paper gives a detailed description of the workflow, which outlines the interaction between the various actors involved in the process of supplying and diffusing results. Finally the integration of the RIS-OSH system with the GL database and the global system architecture are described.

1. RIS-OSH: A system between Current Research Information Systems and Institutional Repositories

This paragraph intends to motivate and describe the choices adopted in the development of the RIS-OSH system. Therefore, it will underline the relationships between RIS-OSH and other system types currently used for gathering and diffusing scientific information, emphasising the referral points taken into account in the design and implementation of our system. Moreover, these relationships show the ever-increasing need [1, 2, 3] to build an integrated environment able to reconstruct the fundamental steps of the scholarly communication chain, which is the major aim of the systems developed in this field.

The design and development of the RIS-OSH is primarily based on Current Research Information Systems (CRIS), as it is properly a "tool dedicated to providing access to and disseminating research information" [4]. However, the new system can be placed in the wider framework of Institutional Repository (IR)¹, both for its features and for its aims, i.e. integrating the different types of information produced within an organisation [5].

Both CRISs and IRs have the purpose of expanding access to the information generated within an institution, thus increasing its visibility and prestige, as well as promoting the impact of its research projects. The main features describing IRs can be therefore applied to CRISs and consequently to the RIS-OSH system [6,7]: their information content is institutionally generated, cumulatively collected, distributedly archived and preserved for future use.

The *relationship of continuity*, which links CRISs with IRs, has been recently underlined by several authors [8], as well as confirmed by the EuroCRIS group [9]. This relationship pertains above all to their common objectives and to their *complementarity* in their information content. CRISs are particularly dedicated to reconstructing the process of research projects carried out within an institution, thus documenting the 'context' in which research outcomes are produced: research community involved, skills and equipment exploited, budget and time employed. This context provides the *background* of a research project, thus providing an essential, *preliminary* part of the IRs' information content: the management and diffusion of digital materials created by the institution and its community members. From the point of view of CRISs, this *digital material* is part of the documentation that constitutes the outcome of a research project.

The design and the implementation of the RIS-OSH system took both CRISs and IRs issues into account. We adopted the CERIF 2000 information model [10], which is a European standard for a comprehensive project description for R&D, thus allowing the interoperability with other CRISs, but we enhanced the data model in order to include digital materials, which are the outcomes of a research project.

CERIF describes the research context in great detail and is flexible enough to be customised to suit local realities, as demonstrated by its application in modelling complex organisations and even developing Corporate Data Repositories [11]. CERIF's weak point is the description and availability of research products and publications, and in particular of the research outcomes represented by GL [12, 13]. RIS-OSH maintains the CERIF distinction between publications and products. However, to solve the cited weakness, it specifies three types of publication: final project deliverables, journal articles and conference presentations, described by their authors, title, data of issuing, and connected references. The system also provides a function for uploading the full-text of the deliverable, which will be stored in the document repository (see § 5).

With regard to products, we considered digital materials such as databases, electronic courseware, prototypes, web pages, and other GL document types such as guidelines, information guides and risk

¹ In the set of services developed within the environment of IRs, we also include the different systems known as e-print archives, since they have been the forerunners of an innovative model for the enhancement of scholarly communication. In fact, IRs' and e-print archives' common feature and strategy are based on open access, which has led to the Budapest Open Access Initiative as well as to the development of the standards supporting interoperability such as the Open Archives Initiative Protocol for Metadata Harvesting.

profiles. These products are described by the title and by a brief abstract as well as giving the URL, where they may be accessible.

Like other IRs, the RIS-OSH system emphasises the role played by the diffusion of digital materials - most of them composed of GL - produced within an institution, considering both the diffusion of their bibliographic description and the availability of their full texts. To facilitate this dissemination, it is necessary to re-construct the organisational model of a project research lifecycle, allowing the different "information producers" to play an active role in the process. This is achieved in our system through a distributed input of the information, which is managed and controlled by the identification of the activities and roles of the different information producers and by authorising their access to the system. This has a twofold advantage: it makes it possible to assess and enhance the quality of information and to speed up the updating of information available both within and outside the institution. For this reason we used workflow methodology in the RIS-OSH system.

2. Workflow function

Workflow technology has been used to re-engineer complex processes within the business and enterprise environments. The conceptual description of the process is used to highlight any bottlenecks that may limit efficiency and to identify the parts of the process that can be automated. The process is also described by identifying the activities and resources (human and equipment), and the times and costs required to perform them. Particular attention is dedicated to the identification of the actors (human resources) that perform the activities, their roles and skills, the interaction between the actors described as an exchange of information and/or as a result of the transactions performed.

The adoption of workflow technology in information and documentation systems such as CRISs [14, 15], IRs [16, 17, 18] and open archives [19] highlights the need to improve the process of collecting and diffusing information within a wider context than the one traditionally performed by the library. In fact, these systems aim to reproduce and enhance the scholarly communication. It is no coincidence that the workflow-based systems primarily collect GL documents or digital materials, that is, documents that are more easily available by their own *information producer*, i.e. the author. For this reason, to improve the process and consequently the system application, it is necessary to reconstruct scholarly communication and this requires a careful analysis of the organisation from both a scientific and organisational point of view relating to the diffusion of results.

In analysing the process of scientific production, which in our specific case is represented by the setting up and performing of research projects, and that of diffusing the results thereof, the following bottlenecks emerged:

- ISPEL is composed of Departments, each one performing research within specific research programs, and adopting its own scientific and organisational procedures for the activation and performance of projects and for the diffusion of results.
- Project results, and in particular the final deliverables tend to remain property of the Department and there is no consolidated procedure for transmitting the documents to other departments and/or the library.
- The Library receives the deliverables resulting from the projects with a considerable delay, and the absence of a standardised format of the deliverables from external structures [20] makes the correct bibliographical treatment of the GL documents more complex.

The development of the RIS-OSH system had, therefore, the primary aim of improving communication between departments, synchronising and shortening timeframes of the various activities connected to preparing and setting up projects and also improving the activities of collecting and diffusing the results. It was, however, necessary to create a new, rather flexible system that, in introducing new work organisation modalities, respected the work methods and characteristics of each Department. This last requirement was satisfied by identifying the actors, for each activity, thus giving the individual departments the responsibility of assigning tasks according to their own organisational modalities.

The process description methodology that characterises a project's lifecycle was dealt with in detail in [21]. It pertains to the set-up process, from the phase of elaboration of a research project proposal for funding, through to its assignment and project conclusion. The same methodology has been now used to describe the integration between RIS-OSH and the GL database. In this case the point of view of the analysis is the exchange of data between different environments, in which each of the various actors plays his/her own role.

Summarising the entire project, we can conceptually point out three sub-processes:

- Entering data concerning the project proposal, including the phases in which the proposals were drafted by each researcher/department; the drafting of the organisation's Activity Plan; the publication of the Call for proposals; and the assignment of the project to an external organisation.

For this sub-process an interface was developed to facilitate data entry, browsing and searching in an intranet environment.

- Collection of project results also includes collecting and archiving data relating to the deliverable, the project's abstract and any indication and description of other products obtained from research. This data is collected through a web interface filled in by the external scientific managers, once the project has been concluded.
- Interaction between departments and the library concerning the data input of the deliverable and the subsequent bibliographic treatment by the library. This latter sub-process takes place automatically, through the integration of the two systems, RIS-OSH and the GL database.

3. Project phases: the process state diagram

A description of the process at a high level of abstraction is shown in the state diagram (fig. 1) that describes the activities and transactions of each phase identified (a more detailed description of the transactions of each phase is given in [22]). The diagram shows that a certain number of states is associated with one particular process phase. The results of the first three phases of the project lifecycle are represented by the production of certain types of document: the Proposal, the Activity Plan, and the Call for Proposals. In the later phases (Project assignment, Conclusion of project), the activities performed are aimed at the completion of information on the project.

The events, that cause the transaction from one state to the other, mainly constitute the positive or negative result of assessment activities performed on the quality control of information (*Positive or negative evaluation*), on its consistency (*data validation*), and on the subsequent permission to diffuse and publish the data to users outside the ISPESL. These activities are performed by different actors. In other cases, the state transaction is given by the introduction of certain information, for example the proposal becomes complete if information is entered in all mandatory fields.

The check and assessment of the data entered, which leads to the publication of the data to external users, guarantees the quality control of data. For instance, the approval of the web publication of the deliverable represents a form of *certification* [23, 24], a type of peer-review, which assesses the deliverable's scientific value as well as the institutional choice on the opportunity to make the full-text accessible to external users.

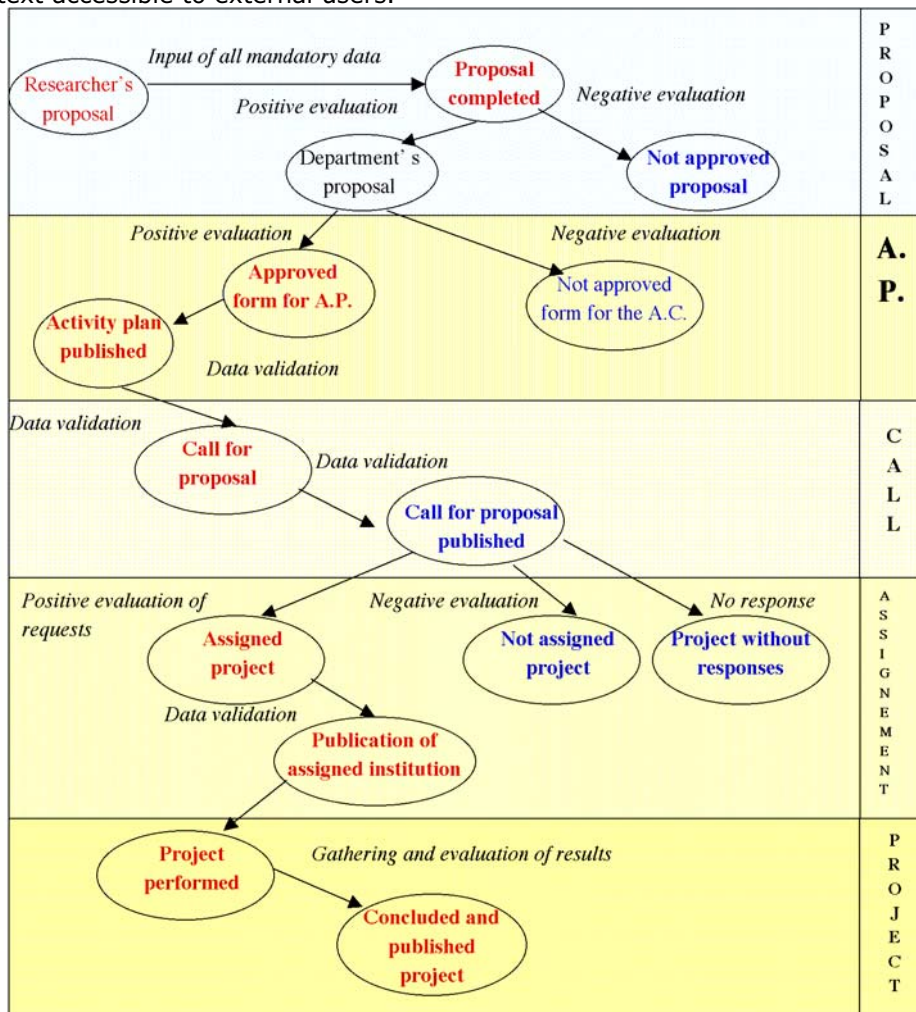


Fig. 1 - State diagram of the project phases

4. Diffusion of results: the interaction between the different players

As previously mentioned, the methodology used to describe the process was also adopted for the collection and diffusion of projects outcomes. In this phase a number of actors interact with one another: the "information producers" and the ISPEL referees and personnel who have the responsibility of evaluating and then diffusing the results.

Figure 2 shows the activities performed by the different actors when the project is concluded. The external authority sends the deliverable to the administration department and this at the same time (parallel activities) fills in a summary sheet containing the required information. The ISPEL referee evaluates both the deliverable and the content of the summary sheet. The sheet is simultaneously sent to the referee and the RIS-OSH manager, who is thus notified that the project has been concluded. Now he/she can ask the ISPEL referee for the result of the project assessment and whether to make the full text of the deliverable accessible. If the ISPEL referee's opinion is positive, the RIS-OSH manager proceeds with the entry of data on the concluded project and with the entry of the full text of the report. Once the information on results has been entered, the automatic procedure described in figure 2 triggers the data input to the GL database, which will then be indexed by the library staff, according to the SIGLE bibliographic rules.

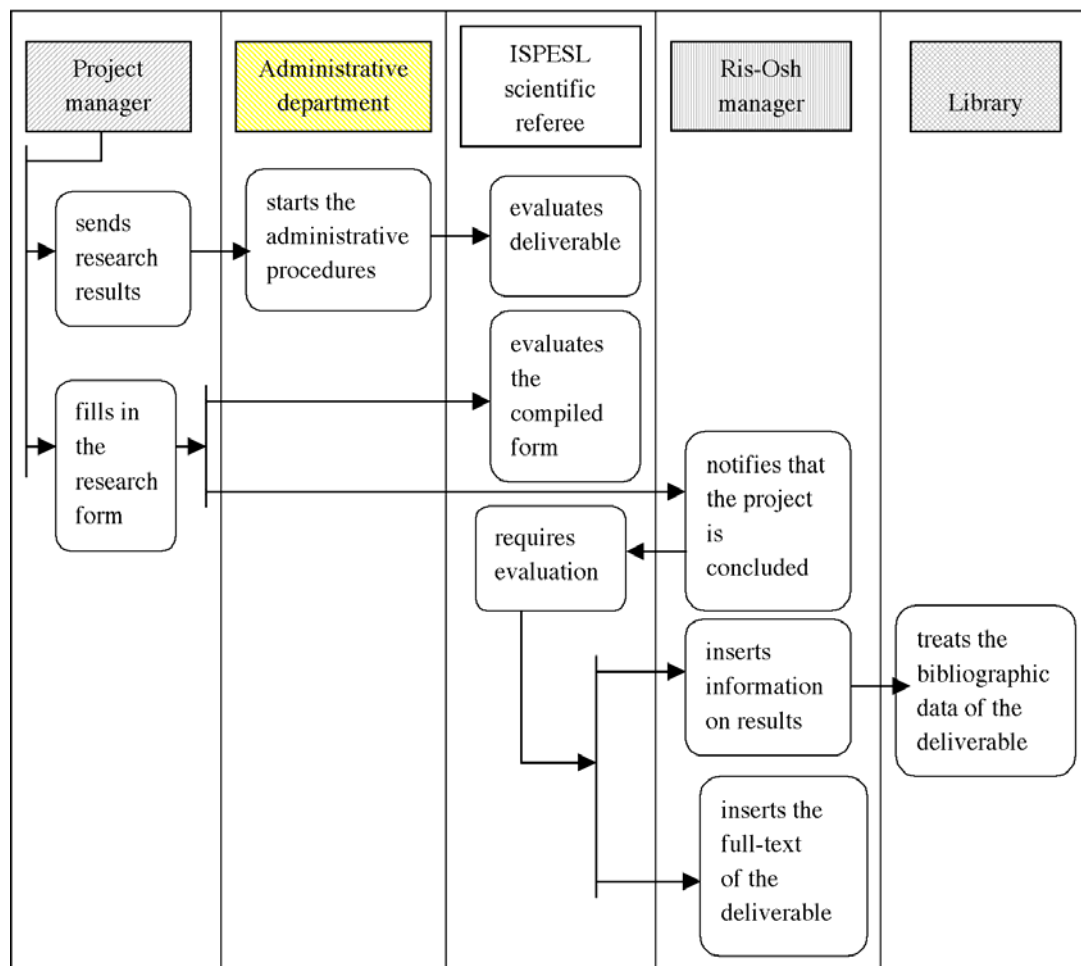


Fig. 2 – The interaction between actors in the diffusion of project results

The automatic procedure of integration between the two systems is in the first place possible from a conceptual point of view, as shown by several studies [25, 26]. It is based on the semantic consistency between certain elements described by the two systems, in particular by the entities *Person* and *Org_Unit* of the CERIF data model. The attribute *role*, present in the CERIF "look-up entities", offers the possibility of describing the entity *Person* identifying the role of the project participants: the ISPEL referee, the project's scientific manager and his/her collaborators, who will become the authors of the deliverable or other research outcomes. Similarly, the organisation's *role* may assume the values of *funding organisation*, *project assignment organisation* and *corporate source* of the deliverable. However, this last value may assume different descriptions, depending on its contexts. In

developing the GL database, particular attention was dedicated to the formulation and standardisation of the corporate source, which, as well known, is one of the most important sources of information retrieval for GL documents. For this reason the SIGLE cataloguing rules were adopted [27]. Furthermore, in order to facilitate the correct description of this important bibliographic element a module of the GL database was developed to facilitate its entry (distinction between primary and secondary corporate source, etc.) and an authority file archive was implemented. However, with regard to a system directly fed by project participants, such a detailed bibliographic treatment, cannot be expected. The aims of this kind of user are rather those of identifying the complete structure, for instance the addresses and contact persons, the descriptions of competence of the participant organisation, that is the information which can help identifying potential partners in other future research projects.

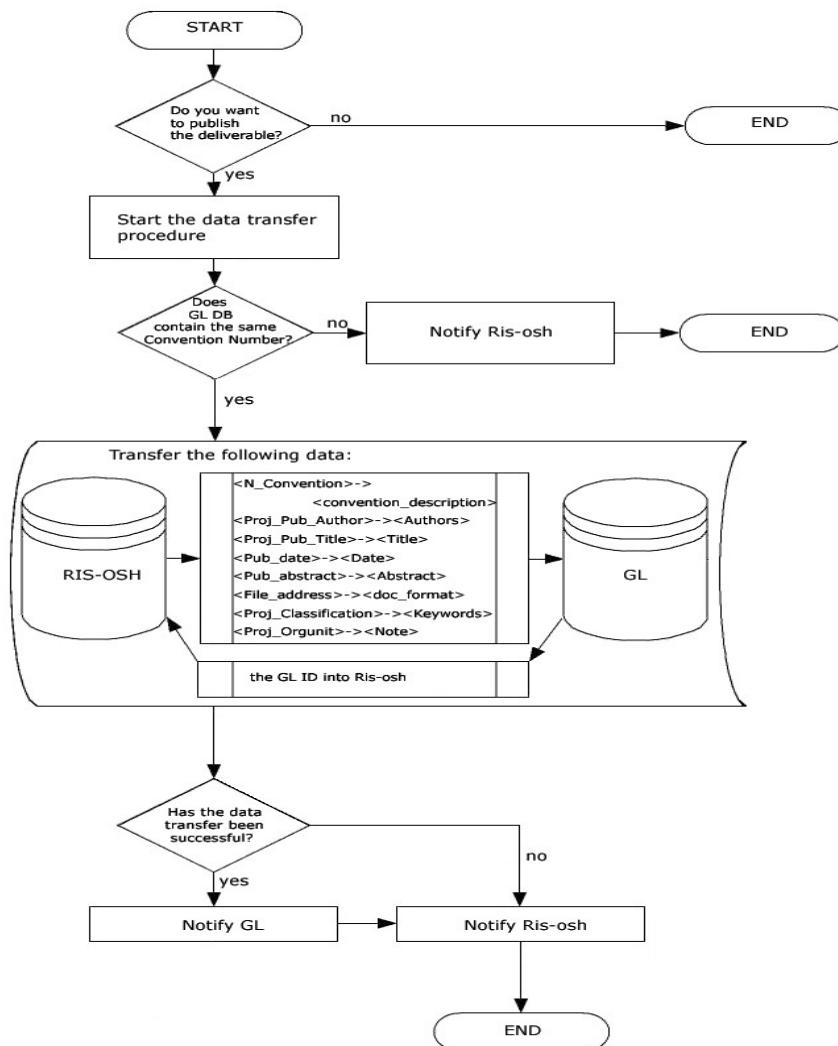


Fig. 3 The flow diagram of RIS-OSH – GL DB interaction

For this reason it was decided to use an automatic procedure for transferring data from RIS-OSH to the GL database, which distinguished between the information that could have been input directly in the GL database fields, from those needing a special bibliographic treatment. Therefore, the transfer of the information related to the corporate source takes place by entering and storing it in a *notes* field, which will subsequently be re-processed according to the SIGLE standards by library personnel. The Library therefore receives all the data on the deliverable as soon as the project is concluded and can proceed with archiving them according to SIGLE standards and if necessary, can access the complete text of the document.

The procedure is described in the flow diagram (fig. 3). It includes two points in which checks are performed: 1) start of the procedure, 2) verification that the GL database does not contain a convention number identical to that given in RIS-OSH (this element has a unique identifier). As soon as this control is completed, the procedure creates a new record in the GL database and subsequently starts the data transfer procedure. The procedure ends notifying the GL database that a new record has been created and notifying the RIS-OSH system of the record identification number.

5. System architecture

The RIS-OSH system is based on an organisational model, in which each department can autonomously assign roles with specific access control. At present, we have identified the following roles: submitter, evaluator, data reviser and publisher. The submitter of information is the researcher, who elaborates the project proposal, and who is generally the ISPEL referee, responsible for the project. However, depending on the internal department's organisation, the information on project proposals can be inserted cumulatively by one of the department's secretaries. The evaluator is the head of a department, who assesses the scientific quality the project proposal; he/she also evaluates, together with the in-house project referee, the quality of the project outcomes and whether to make the full-text of the deliverable available for external users. The other roles belong to the institutional central organisation. The data reviser is responsible for the editorial review of texts, such as the Activity Plan and the Call for proposal, while the publisher is the web manager who can publish the project information on the web for external users, once he/she has received the authorisation.

The main information providers of the system are therefore:

- the researcher/department, which provides for the description of the project proposal and assesses the quality of the research results;
- the external project manager, who performs the project and submits the outcomes to ISPEL's quality assessment;
- the library, which carries out the bibliographic treatment of the deliverable and archives it together with other GL documents contained in the GL database.

Every information provider interacts with the system via a web user interface (back office), which facilitates the submission of data, providing assistance for the correct entry of the information (fig. 4), as well as browsing and search functionalities. Furthermore, the upload of the full-text of the deliverable is managed via the web-interface. An essential component of the system is the workflow, which manages and controls the interaction between the information providers through authentication procedures (*authorisation*) and supports the activities carried out during each phase of the project life cycle (*activity management*). The result of this interaction is a database containing the project description, a repository of full-text deliverable and a *real-time* updated GL database. External users access the RIS-OSH system via a web user interface using browsing and search modules.

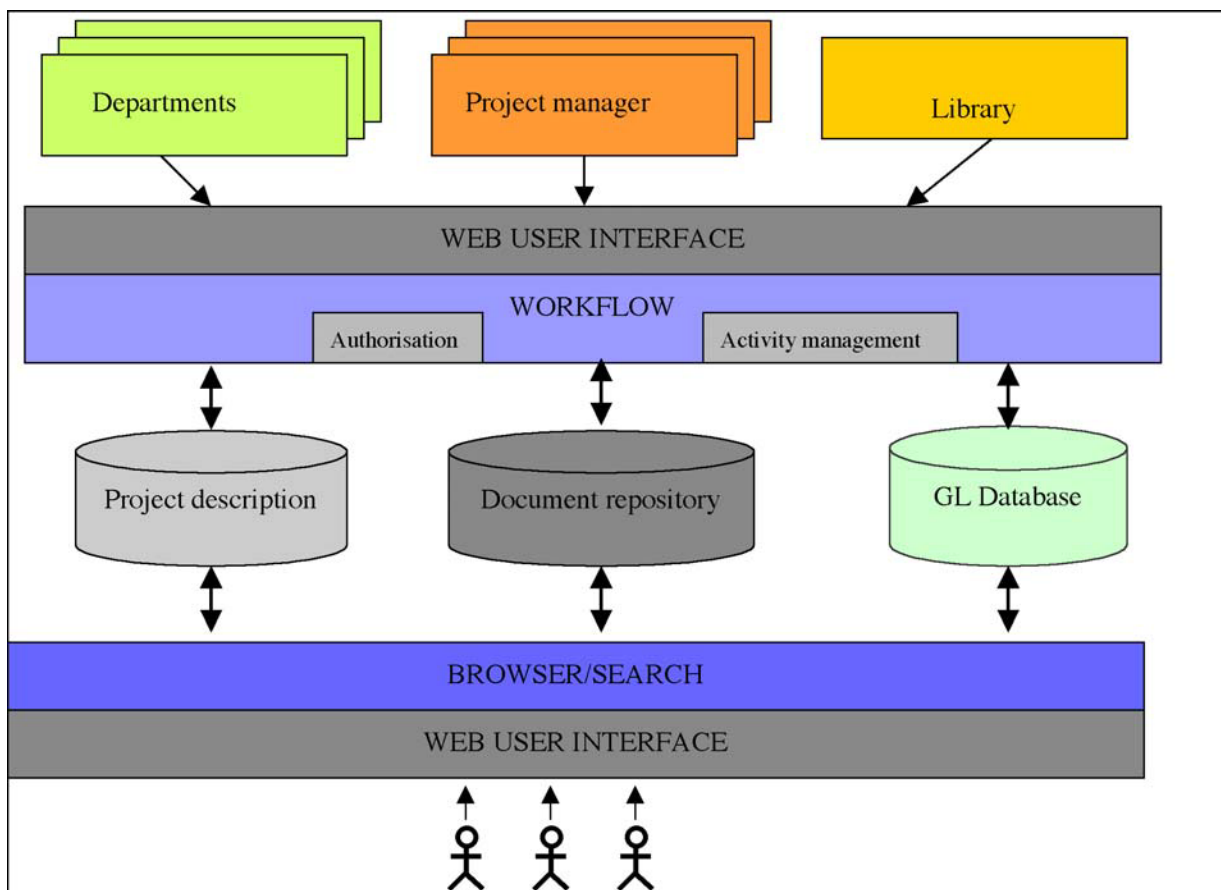


Fig. 4 - The Ris-Osh architecture

Conclusions

The RIS-OSH system will be available on the web interface for external users within the end of this year and it will contain the descriptions of projects carried out from 1999 to 2003. At present, the Information and Documentation Department is testing system functionality related to the processing and publication of next year's Activity Plan and the related Call for proposal.

The RIS-OSH system has been developed giving particular attention to the organisational improvement of the production, collection, preservation and diffusion of institutionally generated information, emphasising the importance of the quality control of the information produced during the project lifecycle. This improvement has been obtained thanks to the reconstruction of the communication chain of the process, under the perspective of the collaboration of the different actors responsible for specific activities. Thus the integration with the GL database, together with the development of an institutional repository, represents the fulfilment of this collaboration, which guarantees the maintenance of a long term archiving of project outcomes.

A final consideration pertains the overall communication chain of research results. It cannot be denied that scientific research is driven, today more than ever, mainly by the ability to communicate the results achieved not only to colleagues and scholars of other disciplines, but also to policy- and decision makers, laymen and to the world of enterprise that constitutes the last link, in which technological transfer takes place.

A Selection of the RIS-OSH system interface

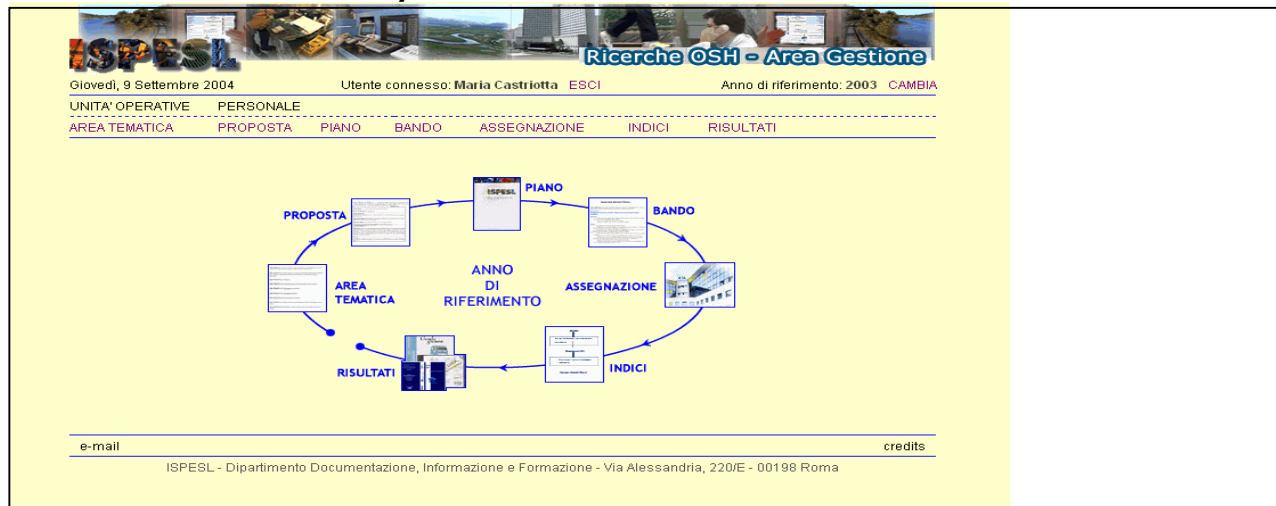


Fig. 5 - The project lifecycle is represented also graphically in the intranet web user interface.

The screenshot displays the RIS-OSH system interface for data entry. The top section shows the title "RISULTATI: inserimento/modifica/aggiorna" and buttons for "visualizza", "chiudi", and "salva e chiudi". Below this, there is a section for project details: "B35/DML/03", "Numero di convenzione: B/35/DML/03", and "Stato: Progetto Concluso (PC)". A section titled "Dati del progetto" contains the following information: "Dipartimento: Medicina del Lavoro", "Titolo: Tutela della salute delle donne della Polizia Municipale", "Referente: Baccolo Tiziana Paola", "Collaborazione alla ricerca: Ente commissionario: Fondazione San Raffaele", "Responsabile scientifico dell'ente commissionario: Paoli Gianpaolo (paolo@rosdsi.it)", and "Data inizio: 23/11/2002". Below this, there is a section for "Data fine" with a date field set to "20/1/2004". A section titled "SEZIONE I: Elaborato Finale" contains a list of sections: "SEZIONE II: Prodotti", "SEZIONE III: Brevetti", "SEZIONE IV: Pubblicazioni", and "SEZIONE V: Note". At the bottom, there is a section titled "Autorizzazione all'aggiornamento della Letteratura Grigia e alla pubblicazione dei risultati" with buttons for "visualizza", "chiudi", and "salva e chiudi".

Fig. 6. Data entry of the projects outcomes

SEZIONE I: Elaborato Finale

Titolo: Con l'espressione "digital born documents" o "digital only" si intendono documenti digitali che sono prodotti esclusivamente in formato elettronico e non hanno un corrispettivo su carta: sono quindi documenti la cui consultazione è strettamente legata all'impiego degli strumenti tecnologici per

Abstract: La crescente produzione di documenti scientifici multimediali di origine digitale, che scaturisce dall'impiego diffuso e diretto da parte dei ricercatori di applicazioni informatiche per la produzione editoriale, ha comportato la necessità di una ricerca finalizzata a sviluppare un approccio innovativo della Biblioteca specializzata dell'Istituto, allo scopo di predisporre ambienti organizzativi e tecnologici per: la raccolta e catalogazione, la conservazione tecnologica e la messa a disposizione in rete all'utenza telematica.
Con l'espressione "digital born documents" o "digital only" si intendono documenti digitali che sono prodotti esclusivamente in formato elettronico e non hanno un corrispettivo su carta: sono quindi documenti la cui consultazione è strettamente legata all'impiego degli strumenti tecnologici per mezzo dei quali essi sono stati creati. Negli sviluppi più recenti della produzione scientifica, si denominano digital born i documenti e artefatti multimediali digitali nell'accezione che li collega al termine, concettualmente più lato, degli exhibit interattivi scientifici (prototipi, dimostratori, modelli funzionanti, ...), con finalità di divulgazione.
Il tema quindi assume una rilevanza strategica, perché introduce sia la visione di nuove forme di produzione intellettuale dei ricercatori, i cui lavori digitali manifestano elevata potenzialità di disseminazione nel web, sia la necessità di preparare il terreno verso soluzioni organizzative e tecnologiche innovative, per coordinare situazioni che ancora non sono gestite nei sistemi informativi correnti.
Dato il carattere marcatamente tecnologico di questi documenti, la ricerca ha inteso contribuire a concepire e studiare un modello innovativo di organizzazione della collezione, per la raccolta,

Autori: -- scrivi QUI gli autori --

Anno: 2004

Documento completo: B35-DML-03-EL.doc
rimuovi ☐
upload Sfoglia...

Contenuti in inglese

Fig. 7 - Data entry for the description of the deliverable and the upload of the full-text document.

Tipo di prodotto: Sito Web

Titolo: -- scegli il tipo --

Descrizione: -- scegli il tipo --

Banca Dati
Guide informative
Linee guida
Pacchetto formativo
Profilo di rischio
Prototipo
Programmi Software
Sito Web

Fig. 8 - Data entry facilitating the choice of the types of project outcomes.

SEZIONE II: Prodotti

Prodotto 1

Tipo di prodotto: Sito Web

Titolo: -- scrivi QUI il titolo --

Descrizione: -- scrivi QUI la descrizione --

URL: -- scrivi QUI l'URL --

Contenuti in inglese

Prodotto 2

Prodotto 3

Fig. 9 - Data entry for the description of other project outcomes

SEZIONE IV: Pubblicazioni

Pubblicazione 1

Tipo di pubblicazione: Relazioni a convegno

Titolo: -- scrivi QUI il titolo --

Autori: -- scrivi QUI gli autori --

Reference: -- scrivi QUI la reference --

Anno: 2004

URL: -- scrivi QUI l'URL --

Contenuti in inglese

Pubblicazione 2

Fig. 10 - Data entry for the description of conventional literature

ISPESL Ricerche OSH - Area Gestione

Venerdì, 10 Settembre 2004 Utente connesso: Maria Castriotta ESCI Anno di riferimento: 2002 CAMBIA

UNITA' OPERATIVE PERSONALE

AREA TEMATICA PROPOSTA PIANO BANDO ASSEGNAZIONE INDICI RISULTATI

PIANO: quadro riepilogativo di verifica per procedere alla pubblicazione

STATO CORRENTE	NUMERO SCHEDE
Scheda Piano di Attività (SPA)	3
Proposta Dipartimento (PD)	0
Scheda Non Approvata nel PdA (SNAPA)	0

STATO CORRENTE	NUM. MANUALE DI CONVENZIONE	NUMERO SCHEDE
SPA	assegnato	3
SPA	non assegnato	0

[procedi alla pubblicazione](#)

PIANO
[convalida](#)
[n. manuale conv.](#)
[visualizza/stampa](#)
[pubblica](#)

e-mail credits

ISPESL - Dipartimento Documentazione, Informazione e Formazione - Via Alessandria, 220/E - 00198 Roma

Fig. 11 - Procedure of the web publication of the Activity Plan

ISPESL ISRDS-CNR		* Gestione Letteratura Grigia *		utente collegato: Maria Castriotta	Giovedì, 16 Settembre 2004
UTENTI DOCUMENTI ENTI PAROLE CHIAVE					
Nuovo Cerca					
Documenti trovati: 16					
Criteri di ricerca - Creato da: ricercheOSH Automatico					
documento 1 di 16					
Visualizzazione del documento n. 229 - tipologia: Relazioni di attività					
Autore/i:	Pippo, A.; pluto, B.				
Ente/i autore:	Attenzione! Campo obbligatorio				
Titolo:	Con l'espressione "digital born documents" o "digital only" si intendono documenti digitali che sono prodotti esclusivamente in formato elettronico e non hanno un corrispettivo su carta: sono quindi documenti la cui co [troncato]				
Anno di pubblicazione:	2003				
Supporto/i:	documento in rete: http://www.ispesl.it/ricercheOSH/risOSH/A2003/B35-DML-03/B35-DML-03-EL.doc				
Convenzione:	Convenzione ISPESL - B/35/DML/03 - 2003				
Collocazione:	Attenzione! Campo obbligatorio				

Fig. 12 - Part of the record automatically created through the integration procedure of RIS-OSH and GL database. The record displayed warns on the bibliographic fields to be completed.

Author Information

Ms. Daniela Luzi
Istituto di Ricerche sulla Popolazione e le Politiche Sociali, IRPPS-CNR
Via Nizza, 128
00198 Rome
Italy

Email: d.luzi@irpps.cnr.it
<http://www.irpps.cnr.it>

References

- [1] Lopatenko Andrei, Asserson Anne, Jeffery Keith G., *CERIF, Information retrieval of Research Information in a distributed heterogeneous environment*, Proc. CRIS 2002 Gaining Insight from Research Information, Aug.29-31, Kassel, <http://www.uni.kassel.de/CRIS2002/programme.gkh> .
- [2] Frey, J.G., De Roure David, Carr Lislie, *Publication at source: Scientific Communication from Publication Web to a Data Grid*, <http://eprints.ecs.soton.ac.uk/archive/00007852> .
- [3] Van de Sompel Hebert, Payette Sandy, Erickson John, Logoze Carl, Warner Simeon, *Rethinking Scholarly Communication. Building the System that Scholars deserve*, D-Lib Magazine, vol. 10, no. 9, Sept. 2004, <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>.
- [4] EuroCRIS, *Transforming Research Information Into Knowledge*, <http://www.eurocris.org/en/eurocris>.
- [5] Crown Raym, *The case for Institutional Repositories, A SPARC Position Paper*, prepared by Raym Crown, 2002, <http://www.arl.org/sparc/IR/ir.html>
- [6] Smith Kackenzie, Barton Mary, Bass Mick, Branschofky Margaret, McClellan Greg, Stuve Dave, Tansley Robert, Walker Julie Harford, *Dspace. An Open Source Dynamic Digital repository*, D-Lib Magazine, vol. 9, no. 1, Nov. 2003, <http://www.dlib.org/dlib/january03/smith/01smith.html>.
- [7] Lynch Clifford A., *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*, ARL Bimonthly Report 226, <http://www.arl.org/newsltr/226/ir.html>.
- [8] Jeffery Keith G., Lopatenko Andrei, Asserson Anne, *Comparative Study of Metadata for Scientific Information: The place of CERIF and Scientific Repositories*, Proc. CRIS 2002 Gaining Insight from Research Information, Aug.29-31, Kassel, <http://www.uni.kassel.de/CRIS2002/programme.gkh> .
- [9] EuroCRIS 2004 Conference, *Putting the Sparkle in the Knowledge Society*, Antwerp, 13-15 May, 2004, http://www.eurocris.org/en/taskgroups/conferences/events/cris2004/_0 .
- [10] Common Research Information Format CERIF 2000, *CERIF 2000 Guidelines, Final Report of the CERIF Revision Working Group Co-ordinated by DG XIII-D-D*, European Commission, <http://www.cordis.lu/cerif>, Sept.1999.
- [11] Grabczewsky Eddy, Crompton Shirley, Robison Stuart, Hall Trudy, *A corporate Data Repository for CCLRL using CERIF*, EuroCRIS 2004 Conference, *Putting the Sparkle in the Knowledge Society*, Antwerp, 13-15 May, 2004, http://www.eurocris.org/en/taskgroups/conferences/events/cris2004/_0
- [12] Jeffery Keith G., Asserson Anne, CRIS, *Grey Literature and the Knowledge Society*, CRIS 2000.
- [13] Asserson Anne, Jeffery Keith G., *Research output Publications and CRIS*, EuroCRIS 2004 Conference, *Putting the Sparkle in the Knowledge Society*, Antwerp, 13-15 May, 2004, http://www.eurocris.org/en/taskgroups/conferences/events/cris2004/_0
- [14] Staab Steffen, Struder Rudi, Sure York, Volz Raphael, *SEAL – a SEmantic portAL with content management functionality*, Proc. CRIS 2002 "Gaining Insight from Research Information", Aug.29-31, Kassel, <http://www.uni-kassel.de/CRIS2002/files/pdf/Staab.pdf>
- [15] Mascord Matthew, Jones Catherine, Lambert Simon, *Experience in Creating an Electronic Publications Archive at a National Scientific Laboratory*, EuroCRIS 2004 Conference, *Putting the Sparkle in the Knowledge Society*, Antwerp, 13-15 May, 2004, http://www.eurocris.org/en/taskgroups/conferences/events/cris2004/_0
- [16] Smith Kackenzie, Barton Mary, Bass Mick, Branschofky Margaret, McClellan Greg, Stuve Dave, Tansley Robert, Walker Julie Harford, *op. cit.*
- [17] Smith Kackenzie, *Dspace for E-print archives*, High Energy Physics Libraries Webzine, no. 9, 2004, <http://libray.cern.ch/HEPLW/9/papers/3/>.
- [18] Beier Gerhard, Velden Theresa, *The e-DOC-Server Project/ Building and Institutional Repository for the Mac Planck Society*, High Energy Physics Libraries Webzine, no. 9, 2004, <http://libray.cern.ch/HEPLW/9/papers/4/>
- [19] Gradmann Stefan, Mruck Katja, Stempfhuber Maximilian, *Refining the Semantics of Open Access*, EuroCRIS 2004 Conference, *Putting the Sparkle in the Knowledge Society*, Antwerp, 13-15 May, 2004, http://www.eurocris.org/en/taskgroups/conferences/events/cris2004/_0
- [20] Aceti Angela, Castriotta Maria, Di Cesare Rosa, Luzi Daniela, *A profile of GL producers in the field of safety and health at workplaces in Italy: results of a sample survey*, GL'99 Proc. On New Frontiers in Grey Literature, Washington, DC, USA, 4-5 Oct. 1999.
- [21] Luzi Daniela, Di Cesare Rosa, Libutti Luciana, Castriotta Maria, *The communication flow of research projects results*, GL5 Proc. Grey Matters in the World of Networked Information, Amsterdam, 4-5 Dec. 2003, also published in: publishing research Quarterly, vol. 20, no. 1, 2004.
- [22] Luzi Daniela, Castriotta Maria, Manco Mariarosaria, Mazzucco Renza, Di Cesare Rosa, Libutti Luciana, *Elaborato finale del progetto annuale: Analisi delle metodologie di trasferimento dei risultati in materia di salute e sicurezza del lavoro, mediante l'uso delle tecnologie di rete*, Roma, CNR-IRPPS, '04.
- [23] Roosendaal Hans E., Geurts Peter A. Th.M., *Forces and functions in scientific communication: an analysis of their interplay*, CRISP 97, Cooperative Research Information Systems in Physics, <http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>
- [24] Van de Sompel Hebert, Payette Sandy, Erickson John, Logoze Carl, Warner Simeon, *op. cit.*
- [25] Asserson Anne, Jeffery Keith G., *op. cit.*
- [26] Luzi Daniela, Di Cesare Rosa, Libutti Luciana, Castriotta Maria, GL5, *op. cit.*
- [27] European Association for Grey Literature Exploitation (EAGLE). *SIGLE Manual* , 3. Ed. Rev/, 1998.

Two Worlds: About Bars and Stars in Scientific Information Publishing, An Analysis of Open Source Ideology As a Means of Self-controlled Publishing

Cees de Blaaij
Library of Zeeland, Academic Department

Scientific publishing has become very profitable for several publishing companies. Information in our age has become a first rate economic asset. Another consequence is that smaller publishing companies have gradually been taken over by bigger ones. Ergo there is less competition and more concentration of economic power in the publishing sector. At the same time it has become more important for the mammoths of the publishing industry to protect these interests and give it a sturdy legal basis. This approach has triggered significant changes in intellectual property laws on a global scale. Global diversity of intellectual property became a global standard because economic powers wanted to control distribution channels to reach customers.

This development did not benefit large groups of authors in general. Especially those authors who had and still have contractually obligations to the bigger publishing companies cannot make their scientific information accessible to the larger public. The simple reason for this is that the price for consumers - like libraries - has become too high. Furthermore that authors have transferred their rights to the publisher. This has resulted in a Catch 22 situation: "you-can-check-out-any-time-you-like but-you never can leave".

In less poetic words: this "for profit" approach has caused an access crisis in scientific information because the ideological and legal basis of the scientific information chain has been disturbed. This article reviews this situation and analyses the viability of present efforts for publishing scientific information (including grey literature) via other kinds of publishing modes based on "open source ideology".

This approach benefits authors of scientific information in general, but especially the authors of grey literature because of the public nature of this type of information. An open source approach counterweights the present economic policies of big publishing houses. The sharing of knowledge is the primary goal based on public interest. Secondly the problem of public access is guaranteed and thirdly the author(s) have more self-determination. They have more control of their situation.

Introduction

The UK House of Commons Select Committee on Science and Technology published a report on science and technology on the 7th July this year¹. In this report the committee stated "there is a mounting concern that the financial benefits from the Governments substantial investment in research is being diverted to an excessive degree into the pockets of publishers shareholders". Ian Gibson, the committee's chairman, even accused commercial publishers even of "ripping off the academic community".²

This statement characterizes for a large part the feelings that exist today about publishers in the intellectual public domain established by the activities of the scientific community. This domain can be considered as a global information pool with two sides fighting for a larger share of the knowledge pool.

The development of new digital technologies opened new ways of distributing and exploiting information. Scientific information became a mining area for commercial publishers: an "intellectual land-grab" was the result. A battleground for the involved scientists and publishers, especially the commercial publishers in the scientific, technical and medical fields.³

The actual setting resembles a Prisoners Dilemma game⁴: "Two men look out the same prison bars; one sees mud and the other stars".⁵ The first player represents the research community for the greater part dissatisfied with the practices of the scientific publishing industry due to control of intellectual property rights and consequential restrictions on the free circulation of scientific knowledge.

The other player represents the publishing industry, trying to exploit the same resource at a maximum profit for his own benefit.

Game theory

Each player is without concern for the well being of the other player. This doesn't lead to an optimum solution. According to the principles of game theory the best strategy for a given player is often one that increases the payoff to one's partner as well. It has also been shown that there is no single "best" strategy; how to maximize one's own payoff depends on the strategy adopted by one's partner. Only by

¹ UK Science and Technology Committee, Tenth Report, Scientific Publications: Free for all?, dd. July 7th 2004; HC 399-I, HC 399-II, source: <http://www.parliament.the-stationery-office.co.uk/pa/cm/cmsctech.htm>

² The Times (London), August 6, 2004.

³ The humanities and social sciences are not considered here.

⁴ A good introduction to Prisoners' Dilemma, including studies of strategies and discussion of the game's significance is Robert Axelrod's **The Evolution of Cooperation** (Basic Books, NY, 1984).

⁵ The Oxford Dictionary of Quotations is the source of this quotation by Frederick Landbridge (1843-1923) from his book, *A Cluster of Quiet Thoughts*, Religious Tract Society, 1896.

cooperation the situation would improve, according to the pay-off matrix⁶. At this moment both players are not working together but a new kind of public-private publishing model is desirable if not vital to serve the needs of the scientific community.

To show this I will first evaluate the present subscriber-pays model of scientific publishing on its impact in respect to the academic community. Secondly I want to deal with the problem: Can a model of common property rights mean – based on what is called Open Access or more specific Open Source – as a model for scientific publishing.

In regard to Open Source we need to take two considerations into account. Besides free access to research, we also need a balanced control of intellectual property rights in the framework of open source ideology. The way grey literature has been handled so far on institutional and subject-specific levels can offer an experience for new ways in Open Access publishing.

Market monopoly

The free market has to a certain degree monopolized the scientific information market.⁷ Reed-Elsevier is the leader in digging the gold mine. Estimates are that they control 20%-25% of sales of the STM market and publish about 1800 titles. This means that they set the tone in the serials marketplace. Behind Elsevier there is group of commercial publishers who resemble the leader in marketing strategy like Springer/Kluwer. This merger was engineered by Candover and Cinven, a venture capital firm specialized in exploiting acquisitions for profit.⁸

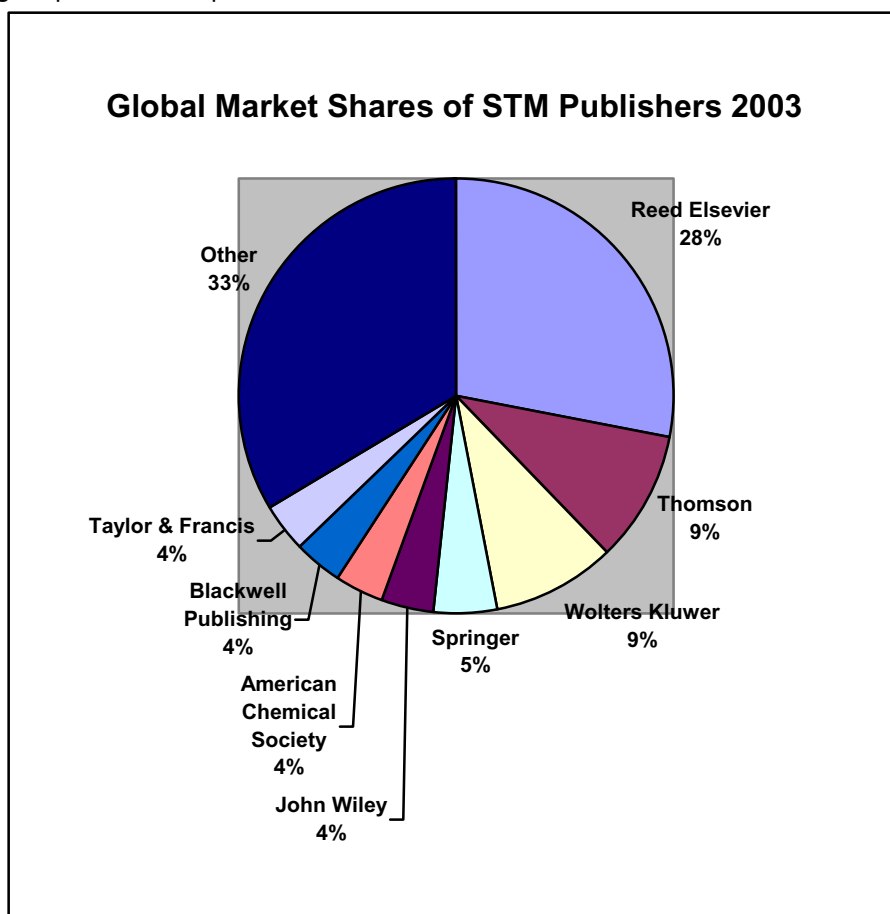


Figure 1.⁹

⁶ A **payoff matrix** or **payoff function** is a concept in game theory, which shows what payoff each player, will receive at the outcome of the game. The payoff for each player will of course depend on the combined actions of all players. For the situation regarding Open Access it resembles a scenario of two states engaged in an arms race. The Open Access Movement and the commercial publishers will reason that they have two options either to use or don't use or come to an agreement. Neither side can be certain that the other one will keep to such an agreement. They both incline towards their own model.

	Cooperate Com. publishers	Defect Com. Publishers
Cooperate OA movement	win-win	lose much-win much
Defect OA movement	win much-lose much	lose-lose

⁷ The situation for the humanities is quite different. Also Open Access works differently Humanities than for STM market, see: Peter Suber, Promoting Open Access in the Humanities, source: <http://www.earlham.edu/~peters/writing/apa.htm>

⁸ CANDOVER/CINVEN/BERTELSMANNSPRINGER (Case No COMP/M.3197-Mergers) [2003] EComm 54 (29 July 2003)

⁹ Source: UK Science and Technology Committee, Tenth Report, Scientific Publications: Free for all?, dd. July 7th 2004; HC 399-I, HC 399-II, source: <http://www.parliament.the-stationery-office.co.uk/pa/cm/cmsctech.htm>, p. 5. The percentages are rounded up or down to whole numbers.

Figure 2 Key figures Elsevier¹⁰

Year	1999	2000	2001	2002	2003
Net Revenues*	2,577	3,091	3,671	3,991	3,571
Net Profit*	-48	27	101	144	242
* in million(s) of Euro					
Dividends					
Fiscal Exercise	2003	2002	2001	2000	1999
Dividend Net (Euro)	0.22	0.21	0.21	0.19	0.15
Earnings per share					
Year	1999	2000	2001	2002	2003
Earnings per share (in Pence)	-0.07	0.04	0.13	0.18	0.31
Sales per activity					
Scientific and medial publications	28.00%				
Legal, fiscal and regulatory publications	27.00%				
Economic information	27.00%				
Education	18.00%				

As we can see scientific publishing is a very rewarding business. The margin of profits has gone up considerably the last four years as a result of ever increasing subscription prices and are as much as 34% for Reed-Elsevier. Elsevier is defending this by saying that such a margin is necessary to ensure the sustainability of STM journals and publishing more papers each year.¹¹ This is not very satisfying if we compare these high profit margins to the general situation of faltering library budgets. In general academic library budgets simply cannot keep up with increased costs and new subscriptions.

Secondly the costs of print publications restrict the amount of information being published. Not many individuals would subscribe to costly journals. So only libraries will buy them as far as this is still possible. In 2003 Cornell University¹² paid \$ 1.7 million for 930 Elsevier subscriptions. This consumed 20% of the university's total periodical budget. That is only 2% of the total number of serial titles to which Cornell subscribes. There was too much financial restraint. The result was massive cancellation of Elsevier titles. Cancellations by other universities like Harvard¹³ and MIT, Duke University, University of North Carolina, North Carolina State University¹⁴ happened for the same reason.

One reason it came so far is that libraries kept on paying to serve the academic community and they relied on high-priced titles with a high esteem for the peer reviewed journals. Publishers believed that libraries would complain, but they would pay up in the end. Stockholders wanted to see rising profits so publishers will raise prices as much they can.

But with the cancellations a new step has been made towards a situation more favourable for Open Access. Cornell instigated after the cancellations they would support the development of Open Access journals as a viable alternative.

¹⁰ Source: Yahoo Finance <http://uk.biz.yahoo.com/p/e/elsn.as.html#2> (consulted 18th Oct. 2004)

¹¹ The Economist, August 5th 2004

¹² Cornell Faculty Senate Resolution. Resolution regarding the Universities Library Policies on Serial Acquisition, with Special Reference to Negotiations with Elsevier. Source: <http://www.library.cornell.edu/scholarlycomm/resolution.html>

¹³ Source: <http://www.news.harvard.edu/gazette/2004/02.05/10-libraries.html>

¹⁴ Source: <http://libraries.mit.edu/about/journals/cancellations.html#access2>

Intellectual property rights

The second problem is the control of intellectual property. There are several implications which are a part of the subscriber-pays model but can be avoided in an open source model.

- Authors provide their article for free to scientific publishers. At the same time they have to transfer their copyright to publishers. They do not get paid, instead they must find acclaim within the academic community. In normal circumstances suppliers of economic goods get paid for commodities.
- Scientific information is becoming more and more centralized due to economic concentration of publishing companies by mergers. This means there is less competition and more restrictions to access related to high prices.
- Publishers limit the use of journals by using access control by IP address recognition, number of simultaneous users etc. Licensing conditions, that do not reflect usage patterns and service requirements in many cases.
- Cancellation of printed journals by libraries can lead to access restrictions of digital back issues due to licensing regime
- The "fair use" principle might be at stake (case Felten vs. RIAA)

Influence of copyright law

A general objective of copyright law is to stimulate the creating of scientific works and make it possible that after a limited time these works become available to the public domain. The introduction of new copyright laws in the late 90's did exactly the opposite. The extension of the duration of copyright protection in Europe¹⁵ in 1993 and the USA in 1998¹⁶ caused the effect that writings belonging to the cultural heritage will stay invisible like unpublished manuscripts needed for research. In the situation of the USA it means that no new works will enter the public domain until January 1, 2019¹⁷. Another example that hampers research is the introduction of new copyright law related to the use of digital technologies.

In the American situation creating excerpts from journal articles for academic purposes is considered to be fair use but the Digital Millennium Copyright Act (DMCA)¹⁸ makes it illegal to create such excerpts by circumventing copy protection on electronic media. Maintaining the "fair use" exception does not provide for enough protection.

The introduction of DMCA brought liability to the scientific world. I refer to the case Felten et. al vs. RIAA (Recording Industry Association of America) et. al.¹⁹.

This was a case in which Princeton academics were suing the recording industry over abuse of the DMCA to silence research that could impinge upon the entertainment industry's desires to protect its intellectual property at all costs.²⁰

The recording industry threatened the researchers under the DMCA for their planned release of a research paper describing the defects in the proposed Secure Digital Music Initiative (SDMI) lock-down schemes for audio CDs. The original threats led the researchers to withdraw the paper from a planned conference. In response to the lawsuit, the recording industry promised not to sue the research team for presenting the research at a security conference in August 2001. Only after getting assurances from the government, the recording industry, and a federal court stated that the threats against his research team were ill conceived and would not be repeated. Felten and his team decided not to appeal. The question is if the current national or international copyright regimes are being efficient – whether it can reach an optimum between private incentives and social benefits. The case of Felten vs. RIAA shows that scientific research can be put in jeopardy. There are quite a number of law cases in which copyright law is supportive in protecting the interests of the entertainment industry and not enough attention is being paid to the issues of freedom of access and of the public domain.

To advance to a more fruitful situation for science and the public domain several requirements have to met:

- Existing copyright law should not be seen as a dogma, but should be used as an instrument to encourage scientific research.
- Placing a low limit on the scope and duration of copyright protection to make sure that scientific information becomes widely available. Scientific information is a strong social asset so it needs to be legally treated as such.

¹⁵ EU Directive on harmonising the term of Copyright Protection, 1993

¹⁶ Sonny Bono Copyright Term Extension Act, October 7th, 1998. The change in the law allows corporations to have exclusive rights for a total of 95 years, instead of 75 years. For individuals, such as authors and songwriters, it extends copyrights to a total of 70 years after death, up from 50 years.

¹⁷ No copyrighted works will enter into public domain due to term expiration in the United States until January 1, 2019,

¹⁸ For an overview see: Digital Millennium Copyright Act: Status and Analysis, source: <http://www.arl.org/info/frn/copy/dmca.html>.

¹⁹ Plaintiffs' Brief in Opposition to RIAA, SMDI And Verance's Motion To Dismissing *Felten v. RIAA* (Aug. 13, 2001), see: http://cyber.law.harvard.edu/openlaw/DVD/cases/Felten/20010813_eff_felten_brief.html.

²⁰ Electronic Frontier Foundation Complaint, *Felten vs. RIAA* (June 6, 2001), see: http://cyber.law.harvard.edu/openlaw/DVD/cases/Felten/20010606_eff_complaint.html. Also other documents at: http://www.eff.org/IP/DMCA/Felten_v_RIAA/ and <http://www.riaa.com/news/newsletter/press2001/042501.asp>

- Revisions to copyright law should not take vested interests or particular interests in question. For a system of property rights to function the cost of establishing and enforcing the regime must not exceed the benefits.
- Too many restrictions on access can ultimately harm the foundations of a democratic society that needs to be well informed.

Subscriber-pays model under review?

What will be the result of all this? Clearly the present subscriber pays model is in crisis. The UK House of Commons' Science and Technology Committee stated that the current model for scientific publishing is unsatisfactory and called on the UK government to support Open Access journals: that is digital, online, free of charge, and free of most copyright and licensing restrictions.²¹ The committee concluded however that the attitude of the government was disappointing and noticed that the British government had "given little thought to the issues so far". The committee hopes that the Report will prove to be a catalyst not only on a national scale but also internationally because of general interest.²²

This hope was not too far fetched as other governmental institutions and universities in the world tackled the same problems.

Just before the UK report was published, the U.S. House of Representatives Committee of Appropriations approved a provision in a bill that backs Open Access to material published by the National Institutes of Health.²³ The committee expressed its concern about the lack of

Public access to research findings and the rising price of journals.

Leading research associations in Germany, France and Switzerland signed the so-called Berlin Declaration – a call for free access to research findings.²⁴

The European Commission has also started an investigation into the publication of academic journals across Europe. The commission is seeking measures to improve access to research. It will look at the trend towards Open Access publishing where researchers pay to publish articles.²⁵ Also the OECD Committee for Scientific and Technological Policy on a ministerial level decided to "work towards the establishment of access regimes for digital research data from public funding"²⁶

Towards an Electronic Open Access Paradigm

A different approach that could fulfil the needs for sharing scientific knowledge and at the same time has a balanced approach towards intellectual property rights is needed. The field of computer science offers comparative solutions. Already in the beginning of the 1980's some computer programmers²⁷ dissatisfied with restrictions to release their programming codes and to build on code of others, started new projects where they had the right to share their work. In 1984 the copyleft mechanism was introduced. Copyleft works under the GNU General Public License²⁸, is a free software license, and gives the right to all users the freedom to redistribute and change software or any program derivated but only if the distribution terms are unchanged. In this way the code and the freedoms become legally inseparable. Copyleft itself is a method of licensing to achieve its goals: to create the most favorable conditions for a wide range of people to feel invited to contribute improvements and/or elaborations to this work, in a continuing process.

In the world of computer science there is a discussion going on which regime of license provides a larger degree of freedom. Basically this is the same discussion as in the scientific community how to provide for Open Access and to promote a freer exchange of information.

The system of scientific publishing requires the attributes of peer review and adequate access control. This working method resembles a model used by software developers known as Open Source.

The Open Source system depends on a developer who is willing to make software code available to the public and uses a system of peer review to test and refine the application.

This leads to debugging and a more stable code.

²¹ For an introduction to Open Access use the Open Access Overview, source: <http://www.earlham.edu/~peters/fos/overview.htm>

²² Ibid

²³ NIH Open Access Plan – Frequently Asked Questions available at <http://www.earlham.edu/~peters/fos/nihfaq.htm>

²⁴ Source: <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

²⁵ The Guardian, City Pages, June 18, 2004, p. 30

²⁶ Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 – Final Communiqué

²⁷ Richard Stallman launched the GNU project in 1983 with the objective of creating a complete free operating system. He is the founder of the Free Software Foundation. To promote GNU, Stallman published the GNU Manifesto in Dobb's Journal of Software Tools Vol. 10, Number 3, March, 1985 " , to bring back the cooperative spirit that prevailed in the computing community in earlier days..."

²⁸ GNU is a recursive acronym for "GNU's Not UNIX"; it is pronounced "guh-noo." For more information: www.gnu.org

The Open Source model provides a developmental model because it also operates on basis of collaboration and peer review. Feller and Fitzgerald stated that the Open Source Model

- Is parallel, rather than linear
- involves mutual collaboration between communities of developers all over the world
- uses independent peer review
- provides feedback to user and developer contributions
- includes participation of motivated developers
- includes increased levels of user involvement
- make use of extremely rapid release schedules²⁹

The Open Source community's use of copyright provides a strong model for the academic community but also from a processual perspective. Science as a social phenomena often used the practice of sharing and reviewing of information. In the Middle Ages literature in the format of theses were nailed to the wall for review. So the academic research model complies with this Open Source ideology.

Another interesting development is also the use of Creative Commons: a legal method which enables copyright holders to transfer some of their rights to the public while keeping others through a mixture of licensing and contract schemes. The target is public domain or open content licensing terms and gives copyright much more flexibility. Present international copyright law does not provide such flexibility for sharing information. The intention is to avoid problems such as I mentioned in the case Felten c.s. Creative Commons was launched in 2001. Movies, books, songs and images are being placed under this umbrella of flexible copyright. Recently in November 2004 Science Commons was launched project to apply the philosophies and activities of Creative Commons in the realm of science. Public Library of Science (PLOS) which has agreed to adopt CC attribution license as its standard license.

Creative Commons underpins also the thought of the 1st US Copyright Act of recreating the balance of copyright and keep copyright duration limited by voluntarily releasing that right after a shorter period.

Another important development is the fact that World Intellectual Property Organization has adopted a development agenda³⁰ that acknowledges the need for balance in the worldwide policy on trademark, copyright and patents. This means attention for open source and the Creative Commons project. The interests of the public had never been represented before at WIPO meetings. With the pressure of non-governmental organisations and adoption of the development agenda the interests of the non-profit are taken more seriously. This is hopeful for the future.³¹

Involvement of Grey Literature

Building collections of grey literature offers good examples of the way it has been made available to the public. Digitalisation and using the Internet for free distributing of information like government information helped the public domain. As the main characteristic of grey literature is its non-commercial nature and supports the idea of Open Access.

It would be interesting to make some cost-benefit analyses in comparing the use of four channels for availability: electronic journals, institutional repositories, subject-specific repositories and self-posting on authors' home pages. These analyses would give us more insight in respect to the viability of the Open Access system.

Subject specific repositories are the oldest. It should be noted that before the coming of the Internet there already existed some tradition in exchanging preprints in research areas where speedy publication was required. Subject repositories allow early and efficient dissemination if contributors involved upload their manuscripts and conference papers in timely manner. Relatively low maintenance costs are involved. Institutional repositories started when universities began to realize that there was a digital information overload syndrome and there was a need to guarantee that research material like theses and working papers would be available for the years to come and systematically maintained.

These repositories also offer an opportunity to avoid the temptation of self-posting. It's important for institutional repositories to make use of common Web standards like the Open Archives protocol.

Self-posting is perhaps the most used method to disseminate information but it is not effective because of a lack of adequate indexing. Using general search engines could be helpful but they hardly can make enough distinction between what is relevant and what is not.

Costs will do it

If Open Access will be a big success depends for a large part on costs and recognition inside the scientific community of the peer reviewed quality of Open Access journals.

On June 17, Credit Suisse First Boston released a report "analysing the effect of the new European inquiry on Reed-Elsevier. From Elsevier's point of view, the good news is that the inquiry is part of a larger plan to double EU spending on scientific research (from 5 to 10 billion Euros/year), which will

²⁹ Joseph Feller, Brian Fitzgerald, *Understanding Open Source Software Development*, London : Addison-Wesley.

³⁰ WIPO General Assembly, Thirty-First (15th Extraordinary) Session Geneva, September 27 to October 5, 2004.

Proposal by Argentina and Brazil for the Establishment of a Developmental Agenda. See:

http://www.wipo.int/documents/en/document/govbody/wo_gb_ga/pdf/wo_ga_31_11.pdf

³¹ Future of the World Intellectual Property Organization, see: <http://www.cptech.org/ip/wipo/futureofwipo.html>

result in many more research articles. Increasing the number of articles published is a venerable justification for journal price increases. The bad news is that the EU already seems to accept that libraries face a pricing crisis, that academics face an access crisis, and that OA is part of the solution. If so, this could endanger Elsevier's 35% profit margins on STM journals. While the EU may not have the power to change the structure of journal publishing in the member countries, the reports from the UK and EU inquiries may stimulate policy changes in the way research is funded in the US, which comprises more than 50% of the STM journal market.³²

The UK-based Wellcome Trust has released a report stating that Open Access could reduce publishing costs by as much as 30 percent. The trust, an independent research-funded charity, estimated publishing costs based on "discussions with individuals in senior positions" at various publishing operations, as well as existing literature on STM publishing.

For a "good- to high-quality journal," the report found, "first copy" costs--the cost of getting an article finalized for publication--would likely be about \$1500 per article under an Open Access model. Overall, the report concluded, the total cost of access to research would add just an additional one percent to the costs of research.

Reed-Elsevier believes the Open Access system is flawed for two reasons: first, it is not covering its costs, and second, it transfers the cost from consumer to producer. Authors and research institutions must pay to be published on Open Access, between \$ 500 to \$ 1,500 a time, but Elsevier believes that the cost of maintaining the scientific system - which requires pieces to be reviewed by the authors' peers in journals - is actually \$ 3,000 to \$ 5,000 a paper.³³ Other publishers have put the costs of article production at £1250, closer to \$2000 per article. Despite the hassle about costs Elsevier has permitted that as of June 3rd authors of papers can post the final text of their articles on the authors' personal or their institutions websites or repositories³⁴ Also Springer will start from January 2005 to put all back issues of it's 1.250 scientific journals online.

So things are moving in the right direction concerning Open Access. Unfortunately a new study³⁵ by the Association of Learned and Professional Society Publishers (ALPSP) shows that 82% of senior researchers (4,000 thousand in 97 countries) knew "nothing" or just "a little" about OA. It makes it clear it will take some time to get scholars acquainted with Open Access.

Author Information

Drs. Cees de Blaaij,
Reference Librarian and Coordinator of Digital Services
Library of Zeeland, Academic Department
Postbus 8004
4330 EA Middelburg
The Netherlands

Eml: C.deBlaaij@zeeuwsebibliotheek.nl
www.zeeuwsebibliotheek.nl

³² SPARC Open Access Newsletter, issue #75, source: <http://www.earlham.edu/~peters/fos/newsletter/07-02-04.htm>

³³ The Independent, February 20, 2004

³⁴ See: http://www.elsevier.com/wps/find/authored_newsitem.cws_home/companynews05_00145 and http://authors.elsevier.com/getting_published.html?dc=PRP

³⁵ ALPSP response to the report of UK Parliamentary Inquiry (1/10/04), source: <http://www.alpsp.org>

The Problems of Grey in the Context of Postmodern Science and Culture

Leonid P. Pavlov, *Deputy Director*
The Scientific and Technical Information Centre of Russia (VNTIC)

Introduction

This paper consists of two parts: the first one where a "theoretical" attempt is made to analyse the general situation in the field of grey literature (GL), and the second part illustrating the practical way medical and healthcare GL is being treated in Russia. The second part is quite classical in its content and intent needs no special explanations or excuses: it is just sharing the Russian experience of medical reports and dissertations collection and dissemination. As far as terminology is concerned, both types of existing in Russia scientific degree works – candidate and doctoral – are termed "dissertation" and can be rendered in English as "thesis" as well so the terms "dissertation" and "thesis" as applied to the Russian scientific scene are used synonymously. The first part falls out of the strictly professional GL field since an attempt to explain a certain grey crisis and keep within the subject area failed.

Part 1. A theoretical reasoning

Recent developments in the realm of grey literature arouse alarming sensations. The capital members of the European Association for Grey Literature Exploitation (EAGLE) abandon the organization, many promising initiatives both national and international (like ISRN) fade, public funding shrinks, commercial projects fail to work, some leading GL experts express their scepticism about the grey future.

At this point I would like to set some limits to the subject of discussion and speak about the "mainstream" grey literature that is GL being a source of scientific and technical information (STI) and among such STI sources speak about scientific research and development reports and candidate and doctoral dissertations – most important and informative types of GL documents I deal with at the Scientific and Technical Information Centre of Russia (the Russian acronym – VNTIC), the federal grey literature center situated in Moscow and functioning since 1967. So, I exclude from my consideration "things like instruction manuals for your toaster, bus schedules, tax forms, etc." which in accordance to Gretta E. Siegel [1] terrify many librarians.

Different ideas have been suggested in order to explain the grey literature situation. Important as they are, the most of the suggestions seem to be of particular, "technical" rather than of principal, fundamental significance and most frequently concern the problems of GL definition, the Internet challenge, a new economic model, blurring the boundaries between the grey and the "white" literature.

Indeed, *the definition* of GL is important but the process of giving a more precise GL specification is endless. The classical definitions by Wood & Smith [2] and that adopted at the 3rd International Conference on Grey Literature (Luxembourg, November 1997) in many respects still work and may serve a working definition at least as far as reports and dissertations are concerned. Grey literature is a fuzzy set rather than a deterministic one and it is difficult if possible to define it strictly. The same is with commercially printed literature. Who can give a unique and monosemantic definition of white literature or fiction? But it doesn't prevent it from being published, sold, bought and read.

Next "much ado" is about *the Internet challenge*. There is no menace to GL in the Internet, on the contrary, the web has enhanced the role of GL. Some authors describe the web as the "new classical grey literature" [3]. The Internet did change our information and communication habits, there is a bulk of GL on the web sites but it doesn't mean that we are supposed to quit our traditional activities of collecting, archiving, and disseminating GL. The Internet is just a tool, a technical and technological instrument that solves some problems and raises others but they are not specifically grey, they have much to do with all sorts of information and documents no matter if they are grey or white.

There are both commercial and non-commercial publishers present on the Internet with the information products both free of charge and liable to charge. Helmut M. Artus [4] concludes that in respect to GL there is "no difference between Internet and the Print". Attributing the Internet challenge to the technical problems he writes: "As to the *technical facilities*, things have widely changed (although printed grey literature is not completely out of fashion), but with regard to the *economic structures* and *social functions* of grey literature we can confirm that only little has changed – if anything at all".

At the same time H.M. Artus goes so far as to suppose that Internet technologies could *blur the boundaries between white and grey literature* and "we could expect, this white literature is going to perish in the long run – and grey literature with it. What remains is only "literature" without any further classification as white or grey". It is hard both to prove and to negate such predictions; it all depends on how "long" this "run" will last. The predictions that the Internet will result in the death of print, white or grey literature reminds that about the death of theatre when cinematograph appeared. Anyway, if finally

grey literature ceases to exist then there will be no subject for grey literature discussions and conferences.

The concern about *new economic models* for grey literature seems more significant since the main GL failures and disappointments mentioned above have been rooted in financial shortages. On the whole, the reaction to the shrinking government financial support is to yield to the “marketplace dictate” and to make grey literature work on a pro-profit basis.

In general most of grey literature community see the way out of the crisis to save the “grey cause” in an adaptation to the new reality. Marie-France Claerebout [5] puts it straight: “Our survival depends on our capacity of adaptation”. This is an evident approach and tactically it may succeed. But it may happen so that in the long run there will be no place for grey literature in a new global paradigm and not because of some possibly natural reasons predicted by H.M. Artus but due to quite unnatural neglect of grey literature. Then what kind of the new paradigm we are supposed to get adapted to?

The answer is *postmodernism* closely connected with economic and political neoliberalism and globalism. The postmodern condition is a global way of thinking and a new ideology that affects grey literature as a part of science, education, culture at large. To understand what is happening in grey literature we have to exceed the bounds of it since its troubles are a consequence of a more general phenomenon. This is like by the second Gödel theorem when it is impossible to prove the non-discrepancy of a system remaining within the limits of the system.

Postmodernism is too extensive and special subject to discuss here in detail still some essentials should be cited basically after the recognized authorities on postmodern – the late French philosopher Jean-Francois Lyotard (1924 – 1998) and the famous Russian politologist, the Solzhenitsyn Prize winner professor Aleksandr S. Panarin (1940 – 2003) who headed the theoretical politology chair at the Philosophy Department of the Moscow State University and to our greatest regret untimely died last September.

At first “postmodernism was a movement in architecture that rejected the modernist passion for the new” and in the early sixties “wanted to maintain elements of modern utility while returning to the reassuring classical forms of the past” [6]. This resulted in a *collage* approach where meaning is found in combinations of already created patterns and “the creative artist was abandoned for playful technician (perhaps, computer hacker) who could retrieve and recombine creations from the past – data alone becomes necessary... We see the dissolution of distinctions, the merging of subject and object, self and other. This is a sarcastic playful parody of western modernity and individual and a radical, anarchist rejection of all attempts to define, reify or re-present the human subject” [ibid.].

Most importantly for our subject is that the postmodern consciousness spreads over not only art and the humanities but applies to politics, economy, science, education and culture on the whole. Denying the central notion of the Enlightenment – the acting subject, postmodernism supports the liberal theory of free market. J.-F. Lyotard [7] puts a fundamental question: what is the role of science in the absence of the big subject of the Enlightenment? In accordance to Lyotard the old social status of science based on the service of quasireligious idols like Progress, People, Freedom, etc. works no longer and the postmodern legitimation of science is market-oriented. The relations between the provider and the user of knowledge will take the value form (*forme Valeur*) and follow the relations between the manufacturer and the consumer of commodity. Knowledge is to be produced in order to be sold.

A.S. Panarin [8] analyses the situation when science is completely reduced to the producer of *knowledge-commodity*. What follows then in respect to science itself and society as a whole? For science itself it is evident: “The market status of knowledge inevitably means the disappearance of those its varieties that failed to take the commodity form and unable to give an immediate (technological) profit. First of all, this concerns basic research, the humanities, all the theoretical education... and is suicidal for science since it loses both the creative imagination and the necessary horizon for the future” [ibid.]. (The translation is mine. – L.P.).

If knowledge is totally reduced to commodity that by definition is open to any buyer then there is a certain probability that *knowledge-commodity* along with all the applied science will be bought by the operators of shadow economy and other shadow practices. Applying this danger to society as a whole A.S. Panarin says that if recently the most shocking interpretation of bolshevism meant lumpens who took over the power to dictate their will to society then now we are supposed to face the perspective when the criminals monopolize knowledge with all the respective consequences to our civilization.

The leading authors of the Fifth International Conference on Grey Literature (GL5, 2003) analyzing the present western situation in scientific information and grey literature, in fact, characterize it as a postmodern condition. Hans E. Roosendaal discussing the information market for research and higher education notes “that research has generally become more subject to market conditions, even when carried out in the environment of a research institution. Market conditions mean that intellectual capital and scarcity of resources, both financial and human, play a more and more important role. As a result, research information is being intensively used for planning and evaluating of entire research programmes emphasizing the formal publication side system rather than the communication side” [9].

The most precise address of grey literature troubles, namely, the postmodern age and its liberal politics, is indicated in the paper by C.Stock and J. Schöpfel. The essence of this politics is the reduction of public research and education funding in every way possible: “Actually, most of our countries are in a difficult economic and financial situation. Governments redefine public priorities, cutting down public

funding, rationalizing and setting for-profit goals to the public sector" [10]. Under the circumstances it is quite expectable that rather ephemeral matters like grey literature are the first to suffer cuttings: "«Marginal» or «fugitive» objects and non-for-profit projects are hard to legitimate" [ibid.]. The results of the liberal politics are revealing: "The French report project completely disappeared, as did the national network for archiving and disseminating of French theses, without being substituted by a workable alternative. From 1998 to 2003, the acquisition of reports, proceedings and dissertations by INIST decreased by 50%, and the French ISRN agency is the only national agency left, without any significant support from the former international agency or the national standards organization. Results of years of continuous and steady work are at risk of disintegration and decay" [ibid.].

The assumption of the global trends prevalent importance supposes a brief comparison between the grey literature situations in Russia and in the West. While the western situation in general and in the sphere of GL in particular developed evolutionary the treatment of grey in Russia changed abruptly with the transformation from the Soviet to post-Soviet reality. What really matters here is that both the evolutionary and the revolutionary ways have led to the same or at least very similar grey situations notwithstanding all the differences between Russia and other scientifically developed countries. The similarity is based on the western neoliberal values adopted in Russia by its ruling elites as a new ideology. The liberal economic model determines new attitudes to scientific research and information services funding.

Russia is often blamed by the West of sins of omission that is being reluctant or too slow to introduce western values. In some respects this may be true. But in fact now we suffer much more trouble because of our sins of commission when we are too fast to implement western patterns completely disregarding the present Russian condition and mentality, the old Russian traditions and experience. The neoliberal politics in science, education and culture is a notorious example of such haste. The government is stubborn to press on the western structures of science and education which resulted in almost ruining the excellent existed system and in creating next to nothing. The latest published government conceptions of reforming science management and higher education tailored after the western models aroused unanimous rejection by scientific and academic community and all the people of common sense.

The Soviet period of Russian history has been comprehensively and broadly criticized and condemned and nobody in Russia wants to loose the achievements of freedom gained in the last fifteen years. However, it would be wrong to deny successes achieved in science, education and culture in the Soviet Union, no matter that the Soviet system proved to be undemocratic. It is known from history that economic upsurge or democratic system does not guarantee spiritual development, creative initiative and the flourishing of science and art. There is a scene in the 1949 movie "The Third Man" in which the cynical character played by Orson Welles says: "In Italy for 30 years under the Borgias they had warfare, terror, murder, bloodshed – they produced Michelangelo, Leonardo da Vinci and the Renaissance. In Switzerland they had brotherly love, five hundred years of democracy and peace, and what did that produce? The cuckoo clock" [11].

The grey literature collection and dissemination system still works in Russia mainly due to the stable foundation laid in the late sixties of the last century. Now the pressing problem is rather of survival than development. Closing this "theoretical" part I conclude that it is not the market paradigm of neoliberalism but the proper state funding that will secure the development of grey literature. If grey community fails to argue governments into the necessity of such funding the "grey cause" will be lost. Dr. H.M. Artus is quite right claiming that "the scientific system" assumes "the political system which is responsible for research funding and support of research institutions and the like" [4] meaning that "the like" should include grey literature centres and libraries.

Part 2. A practical example

In this part a practical example is considered to illustrate how GL is treated in Russia of today, mainly from force of inertia on the basis of the system and infrastructure established in Soviet times. Of course, it doesn't mean that the system stays on the same technical level – most modern digital and network technologies are introduced but organizationally and structurally the system is the same.

A well-developed infrastructure for scientific and technical information was established in the Soviet Union in the late 1960s – early 1980s. This infrastructure laid down the foundation for the so-called State System for STI (the Russian acronym – GSNTI). Among many information centers and libraries – members of GSNTI – VNTIC is a special information centre of a federal level with the aim of collecting and controlling Russian grey literature. VNTIC is responsible for the maintenance of the complete repository for scientific research and development reports and dissertations and carries out the dual function of both archiving and disseminating GL. A more detailed outline of GSNTI and VNTIC activities is given elsewhere [12]. For the sake of example with some quantitative data and qualitative estimations let us consider how Russian GL on medicine and healthcare is being collected and controlled.

There are about 1500 medical libraries in Russia dealing chiefly with white literature. The State Central Scientific Medical Library situated in Moscow supports a complete collection of candidate and

doctoral dissertations on medicine and healthcare stored on paper. There are some other less important branch-specific information bodies supporting partial GL collections and databases on medical conferences, equipment and instruments, statistical data and reports, medicines and drugs, sanitation and hygiene, etc.

In accordance with the Federal Law of the Russian Federation "On the obligatory copy of documents" and the Statute of GSNTI scientific R&D reports and dissertations (candidate and doctoral) covering all areas of human knowledge arrive at VNTIC (about 10 thousand reports and 25 thousand dissertations a year) and stored full-text and in abstract database structure to facilitate information search and retrieval and abstract journals publishing. The total amount of full-text and abstract GL documents in the VNTIC collection exceeds 7 million items.

Table 1 shows the distribution of documents by subject categories in the VNTIC collection in 1996 and in 2001.

Table 1. Distribution of documents by subject categories

	1996	1996	2001	2001
	Dissertations (%)	R&D reports (%)	Dissertations (%)	R&D reports (%)
Social sciences and the humanities (sociology, history, economics, education, etc.)	15	13	43	13
Natural and exact sciences (mathematics, physics, biology, chemistry, geology, etc.)	30	9	16	24
Technical and applied sciences (mechanical engineering, chemical technology, agriculture, medicine and healthcare , etc.)	52	76	39	58
Interdisciplinary problems (space exploration, ecology, etc.)	3	2	2	5

There is an evident decline in technical and applied sciences research (as reflected both in dissertations and reports) and a significant growth of dissertations in social sciences and the humanities. Technical and applied sciences need most expensive research so they suffered badly from the dramatic decrease of the scientific research budget. Besides, the decline is explained by the fact that manufacturing industry and agriculture fell into decay in Russia and there is no demand for scientific research from industrial and agricultural corporations. The growth in the number of social sciences dissertations also doesn't make us happy because since early nineties lots of higher education institutions, mostly private and conferring degrees chiefly in economics and law (the disciplines that need no expensive laboratory equipment and experimentation) have been mushrooming all over Russia but the quality of the education leaves much to be desired.

The subject classification of reports and dissertations is carried out at VNTIC using The State Subject Classifier for Scientific and Technical Information that is a standard 3-level subject classification scheme for science and the humanities in Russia. There are more than 80 first-level subject categories in it. Medical grey literature is classified within first-level subject category 76 – *Medicine and Healthcare* (M&H) which contains 9 second-level subject headings (general problems of M&H; medical biology disciplines; medical materials, means and goods; medical engineering; clinical medicine; clinical pharmacology, pharmacy; hygiene and epidemiology; other fields of M&H; social hygiene, healthcare organization and management) and 134 third-level categories.

There is a second classification scheme used for dissertations – Scientific Speciality Classifier issued by The State Higher Attestation Committee, the official body that confers scientific degrees in Russia. The Classifier contains nearly 5 hundred 3-level codes for specialities in science and the humanities. Code 14.00.00 designates *Medical Sciences* and unites 44 medical professions.

In Table 2 the total amount and the number of medical R&D reports and dissertations (candidate and doctoral) presented in Russia in 1982 – 2003 and in 2003 are shown. It is seen that the volume of medical research has lately increased as compared to the average figures for 22 years (from 5,4% to 9,6% for reports and from 15,6% to 18,8% for dissertations). For dissertations category 76 and code 14 practically overlap but subject heading 76 evidently covers a wider spectre of subjects than speciality code 14.

Table 2. R&D reports and dissertations: total amount and on medicine

	1982 – 2003	2003
R&D reports, total	1220267	13015
R&D reports on Medicine & Healthcare (subject heading 76)	65595	1250
R&D reports on M&H, % of total	5,4	9,6
Dissertations, total	443259	20810
Dissertations on Medicine & Healthcare (subject heading 76)	73075	3910
Dissertations on M&H, % of total	16,5	18,8
Dissertations on Medical Sciences (speciality code 14)	69309	3094
Dissertations on Medical Sciences, % of total	15,6	16,5

In Table 3 a more detailed section for 2003 Medicine and Healthcare (subject category 76) reports and dissertations by the second-level subject headings is given. 73% of medical dissertations and 54,3% of medical reports fall within category 76.29 – *Clinical medicine* that mostly coincides with the subjects of speciality code 14.

A certain part of abstract documents for reports and dissertations is translated into English to be presented in SIGLE database, on VNTIC web-site and published as the English language abstract journals. Table 4 gives some figures for report and dissertation abstracts in English translation for 10 years (1991 – 2001). The English abstracts on Medicine and Healthcare make 4,4% for reports and 5,7% for dissertations.

Table 3. R&D reports and dissertations 2003 on Medicine and Healthcare: second-level subject headings

	Reports	% of total on M&H (1250)	Dissertations	% of total on M&H (3910)
76.01 General problems of M&H	38	3,0	40	1,1
76.03 Medical biology disciplines	101	8,1	246	6,3
76.09 Medical materials and goods	37	3,0	23	0,6
76.13 Medical engineering	130	10,4	38	1,0
76.29 Clinical medicine	679	54,3	2855	73,0
76.31 Clinical pharmacology	30	2,4	124	3,2
76.33 Hygiene and epidemiology	143	11,4	234	6,0
76.35 Other fields of M&H	42	3,4	169	4,3
76.75 Social hygiene	50	4,0	181	4,6

Table 4. Report and dissertation abstracts in English translation (1991-2001): total amount and on Medicine and Healthcare

Report abstracts in English, total	30809	Dissertation abstracts in English, total	41119
Report abstracts in English on M&H	1354	Dissertation abstracts in English on M&H	2341
M&H % of total	4,4	M&H % of total	5,7

The same or more detailed way different scientific subjects can be analysed using the data from the VNTIC grey literature collection that reflects the Russian scientific condition in recent times and at present.

Concluding remarks

In short, three points should be mentioned at the end of this paper.

1. Scientific grey literature processing, collecting and archiving should be government funded despite the antiscientific and anticultural postmodern market paradigm.
2. Grey literature system still works in Russia mainly due to the foundations laid in Soviet times and in spite of the present neoliberal politics.
3. VNTIC supports the most complete scientific grey literature collection on medicine and healthcare in Russia for the latest 25 years.

References

1. Siegel, Gretta E. Capturing Academic Grey Literature – Starting at Home// GL5 Conference Proceedings. 5 Internat. Conf. On Grey Literature, Amsterdam 4-5 Dec. 2003/ GreyNet. – Amsterdam: TextRelease, Jan. 2004.
2. Wood, D.N. and Smith, A.W. SIGLE: A Model for International Co-operation// Interlending & Document Supply. – 1993. Vol.21, No 1. Pp. 18 – 22.
3. Gelfand, J. Grey Literature in New Packages: Implications from the Transition to Electronic Publishing and Information Seeking// GL'95 Conference Proceedings. 2 Internat. Conf. On Grey Literature, Washington D.C. 2-3 Nov. 1995/ Amsterdam: TransAtlantic/GreyNet GL Network Service. 1996.- pp.18 – 32.
4. Artus, Helmut M. Old WWWine in New Bottles? Developments in electronic information and communication: structural change and functional inertia// GL5 Conference Proceedings. 5 Internat. Conf. On Grey Literature, Amsterdam 4-5 Dec. 2003/ GreyNet. – Amsterdam: TextRelease, Jan. 2004.
5. Claerebout, Marie-France. Grisemine, a digital library of grey university literature// GL5 Conference Proceedings. 5 Internat. Conf. On Grey Literature, Amsterdam 4-5 Dec. 2003/ GreyNet. – Amsterdam: TextRelease, Jan. 2004.
6. Morley, James. Postmodern Culture. <http://jefferson.village.virginia.edu/elab/hfl0242.html> (visited 30.09.2004).
7. Lyotard, J.-F. The Postmodern Condition. SPb.- 1998 (In Russian).
8. Panarin, A.S. The Globalism Temptation. Moscow.- Algorithm. – 2001. 558 pp. (In Russian).
9. Roosendaal, Hans E. The Information Market for Research and Higher Education. How to integrate all relevant information in a network of repositories? // GL5 Conference Proceedings. 5 Internat. Conf. On Grey Literature, Amsterdam 4-5 Dec. 2003/ GreyNet. – Amsterdam: TextRelease, Jan. 2004.
10. Stock, C. and Schöpfel, J. Grey Literature in an Open Context: From Certainty to New Challenges // GL5 Conference Proceedings. 5 Internat. Conf. On Grey Literature, Amsterdam 4-5 Dec. 2003/ GreyNet. – Amsterdam: TextRelease, Jan. 2004.
11. Barber, Tony. Switzerland Reconsiders Its Neutrality. The Moscow Times. October 27, 1994.
12. Pavlov, Leonid P. The state and development of the Russian grey literature collection and dissemination centre// Interlending & Document Supply. – 1998. Vol. 26, No 4. – Pp. 168 – 170.

Author Information

Dr. Leonid P. Pavlov, Deputy Director
The Scientific and Technical Information Centre of Russia (VNTIC)
14 Smolnaya St.
Moscow, 125993
Russia

Email: pavlov@vntic.org.ru
URL: www.vntic.org.ru

Towards a Genealogy of Grey Literature via Newton's Journals

Hal von Hofe
CYBEK

ABSTRACT

The question "Is Grey blind to the Review Process?" suggests a number of contradictory answers, all somewhat begging the question of how we identify Grey Literature. In this study the problem will be approached from an historical perspective.

A mass of grey literature from the pen of Isaac Newton was rediscovered in the past century. In 1936 the economist John M Keynes bought a lot of papers at auction at Sotheby's, among which he discovered the so-called "alchemical journals" of Isaac Newton, voluminous working papers devoted largely to Newton's alchemical and code experiments, which he pursued throughout his life. Only in the past 30 years or so have biographies of Newton appeared which attempt to deal with this 'grey' question.

Newton's journals themselves base themselves on a body of grey literature scholars have recently begun to discern under the general notion of "Hermetic writings", which date back at least to the building of the first University at Alexandria by the Ptolemies (3rd century BC Egypt) and generally claim a much greater antiquity than that.

Newton's own description of the reasons for keeping alchemical and hermetic writings away from public review (keeping them grey, as it were) can be found in a letter he wrote to a fellow alchemist, Robert Boyle:

"Because the way by which the Mercurial principle may be impregnated has been thought fit to be concealed by others that have known it, and therefore may possibly be an inlet to something more noble that is not to be communicated without immense damage to the world if there be any verity in Hermetic writers. There are other things besides the transmutation of metals which none but they understand."

On the level of peer review, though Newton has become a sort of icon of the mechanistic universe, both Leibniz and Descartes expressed some suspicion of his notion of gravity, as action at a distance, something that smacked a bit too much of medievalism and magic. (And indeed, the problem of gravity remains a bit of a sticking point in physics today, on its way to the Holy Grail of a unified field theory, caught up in relativity and quantum chaos.)

It is expected that a careful review of the problematic of Newton's 'grey' versus his 'review processed' work, considered against the backdrop of the larger question of 'hermetic writings' in general, will reveal the nature of the link between such literature and the development of mathematics and science in history. It will at the same time permit us to gain a clearer definition of grey literature from the perspective of one of its oldest and most persistent forms.

The question "Is Grey blind to the Review Process?" is here addressed from an historical perspective, taking the case of Isaac Newton as exemplum.

Emphasizing the link between an emerging awareness of grey literature and cybernetic data processing – I believe the term was coined back at the dawn of the cybernetic age in the Kennedy administration – I must begin by acknowledging my indebtedness to the work of the Newton Project, Imperial College London, which, with the aid of the document management system developed by the Australian Science, Technology and Heritage Centre (AUSTEHC), Melbourne University, has provided, on the web --<http://www.newtonproject.ic.ac.uk/index.html> -- an indispensable resource, replete with bibliography, locations and descriptions of extant Newton manuscripts, etc., without which my researches would have been much more limited.

Aside from his published (review processed) 'scientific' writings, Newton's literary remains, his grey literature, may be generally broken down into four divisions: scientific and mathematical papers, theological and chronological writings, writings having to do with his duties as Warden and then Master of the Mint in England, and the alchemical papers.

Some of the scientific papers appeared here and there in drips and drabs after Newton's death, along with a few of the theological/chronological papers, such as his "Chronology of Ancient Kingdoms Amended" and the well known 'An historical account of two notable corruptions of Scripture in a Letter to a Friend' (25,000 words, sent in two parts as a letter to John Locke), to which we shall return later.

Newton's various manuscripts passed through a number of hands. After Newton died intestate in 1727, his friend John Conduitt posted a bond for his debts and claimed the bulk of them. Conduitt expended some scholarly effort with them, but very little saw publication. Over the course of the following century, many of the papers found their way into university collections, but much of the material became scattered and confused along the way.

Bishop Samuel Horsley edited a major collection of Newton's "Complete Works" in 1779-85, which served as a standard for over a century. This collection, however, avoided Newton's alchemical interests almost entirely and suppressed any texts which showed the 'heretical' nature of many of Newton's religious beliefs -- notably the 'third letter' to Locke (following the "Two notable corruptions" essay), in which the unorthodox nature of Newton's belief was more clearly revealed.

In a return of the repressed, however, a mass of this grey literature from the pen of Isaac Newton came to light again in the past century. In 1936 the economist John Maynard Keynes bought a lot of papers at auction at Sotheby's, among which he discovered the so-called "alchemical journals" of Isaac Newton, the voluminous working papers devoted largely to Newton's alchemy. Upon this discovery, he went about tracking down more of these papers and buying them up when he could. He soon found himself in competition with a Jewish scholar and businessman, Abraham Yahuda, who was himself particularly interested in Newton's theological material. Keynes chose to concentrate on the alchemical, and the two worked out an agreement whereby they occasionally exchanged papers they had collected, theological writings from Keynes for alchemical from Yahuda. Keynes collection went to King's College, Cambridge upon his death in 1946. The bulk of Yahuda's collection went to the Jewish National and University Library at Jerusalem, and has been generally available to scholars since the 1960's. Only in the past 30 years or so have biographies of Newton begun to attempt to deal seriously with this large 'grey' mass that has re-emerged into the light of day after having been carefully forgotten over a period of two centuries.

Here I stop to ask if it is grey literature that is blind to the review process, or if it is the review process that tends to be willfully blind to grey? I suspect it is the latter (– though that seems to be changing now, as evidenced by conferences such as the present one). The review process sets as its goal an idealized literature – in the case of Newton this idealization involved making him a figure of scientific genius, a founding father of modern day science. A bit of his theological writing passed review, enough to show a scientist giving the nod to religion, but the writings which showed him to be somewhat heterodox and even heretical in his views did not. The alchemical writing presented a greater difficulty, and was suppressed almost in its entirety.

Its recent re-emergence presents scholars of science and of Newton with a vexing problem. The two most popular explanations for it, both unsatisfactory, are 1) that it really just involved chemistry experiments in nascent form, or 2) that Newton was a man of two worlds, the modern scientific one and the old superstitious medieval one. #2 does not really explain anything, but rather just presents a crack in the idealized figure of Newton as scientist. #1, though it may in some sense be true of a good bit of Newton's alchemical work in a partial sense, does not stand up to comparison with Newton's actual writings on the subject, which regularly imply that it was a good bit more than that.

Newton himself tended to keep his alchemical work out of the public eye and away from any 'review process'. In a letter he wrote to a fellow alchemist, Robert Boyle, he gives the following reason for keeping it grey:

"Because the way by which the Mercurial principle may be impregnated has been thought fit to be concealed by others that have known it, and therefore may possibly be an inlet to something more noble that is not to be communicated without immense damage to the world if there be any verity in Hermetic writers. There are other things besides the transmutation of metals which none but they understand."

What is this danger of "immense damage to the world"? What are these "other things besides the transmutation of metals"? He certainly does not seem to be speaking of simple chemistry here.

As a young man we have evidence that Newton used to record his various 'sins', with a view to purifying himself of such. A brief perusal of alchemical literature reveals this very theme – purification of sin and corruption into higher and nobler states – as a regular topos. A recurrently drawn distinction is made between 'vile gold' and 'the nobler gold', whose purification is said to occur in the heart of the alchemist himself.

It is this last that perhaps most disturbs those who would see Newton as a modern scientist. In the modern view, a scientist is not supposed to experiment on himself, should remove himself as much as possible from his experiments, should be objective, should keep his personal biases, and actual life, out of the science he does. Newton's alchemical work seems to violate this modern proscription, whereby the scientist is not to apply science to oneself.

It is perhaps here that the communication that threatens the "danger of immense damage to the world" which Newton fears can be found, as well as an explanation for the relatively late development of the so-called 'soft sciences', from biology on up to psychology and sociology. I do not think it an accident that Newton's concealed lifelong flirtation with such danger finds itself linked to similarly concealed suspect religious beliefs. A real science of the self, a science of the real, *is* immensely threatening to the established order of the world, society, and the self.

Here we can begin to see why Grey Literature of all kinds, as a repository of potential or hidden truth, is often engaged in resistance with the 'review process' – from the case of Newton's 'heresy' (his theological writings) and 'superstition' (his alchemical and hermetic writings) to that of potentially scandalous governmental memos today. (Imagine for example certain grey documents emerging that detailed a correspondence between American Vice President Cheney and the Halliburton company – we can expect the coming to light of such (to the eyes of a review process) to be heavily resisted, somewhat in the manner of President Nixon earlier, with his claim of 'executive privilege' prior to his resignation -- a claim to copyright over said literature, as it were.) -- Of course, it is the promise of this same potential or hidden truth to be found in grey literature – whether it is actually there or not – that generates its attraction to review as well as its resistance.

Paralleling the case of Newton's own grey writings is the further body of grey literature on which his alchemical work bases itself and to which it contributes, that produced by the "Hermetic writers" he mentions in his letter above. Until recently it too has been largely ignored or dismissed by scholars and the academic review process. Under the influence of scholars like the late Frances Yates (author of *Giordano Bruno and the Hermetic Tradition*, *The Rosicrucian Enlightenment*, and *The Art of Memory*) and others following in her footsteps, this is now changing, and a picture of this 'hermetic tradition' is now emerging that shows it as a running grey trace of the history of the development of science itself.

This literature, which can be identified under the general heading of "Hermetica", dates back at least to the building of the first University at Alexandria by the Ptolemies (3rd century BC Egypt) and generally claims a much greater antiquity than that. Its recognized foundational text is the so-called *Corpus Hermeticum*, a set of dialogues between 'Hermes Thrice-Great' and his various sons, Tat, Asclepius, and Amon, which was most likely written down in Greek somewhere around the time of Christ. Hermes – Egyptian Thoth – is the god associated in mythology with the development of writing and the sciences. In hermetic literature the figure of 'Thrice Great' Hermes is presented as an historical personage who actually existed in the flesh. One can, for example, find references to this Hermes having been an Egyptian King who built a temple to the Sun at Thebes, and later a whole city to the sun elsewhere in Egypt. Though one can find this reference in hermetic literature dating back over a thousand years (-- in the so-called *Picatrix*, originally the Arabic "Ghayat al Hakim", to cite a text that enjoyed a certain popularity in the European Renaissance --) it is only a little over a century ago that we rediscovered archeological evidence of an Egyptian King who had done just that: Akhenaton, husband of the famous Nefertiti (– movie coming soon to a theater near you, by the way!...) Much was also made in the hermetic literature of the Renaissance of said Hermes Trismegistus having been historically contemporaneous with Moses and Orpheus -- something Newton would have been well aware of: he had himself made a translation of the notorious "Emerald Tablet of Hermes", providing a further link to his interest in theology, mythology, and biblical chronology.

In Alexandrian (Hellenistic) times, which saw the emergence of hermetic literature, seven basic sciences were recognized: grammar, rhetoric, dialectic, arithmetic, geometry, music and astronomy. Several centuries before this we find the seminal figure of Pythagoras, with his famous theorem and Orphic cult of mathematical harmonics. Though Pythagoras first formulated the theorem in question, it was already in practical use when the Egyptians built the first pyramids, and the Greeks in general looked back to the Egyptians as the primary source of knowledge upon which their sciences were based.

With the rise of the Roman Empire, this Greek knowledge and science spread further, up until about the 5th century AD, when a certain 'review process' began to suppress it, with the result that the bulk of Greek knowledge was lost in the 'West'. It did, however, continue to flourish in the Arabic world during the early centuries of Islamic culture. A major contributor to this further development of Greek science in the Arabic world was an Hermetic community centered in the city of Harran. The so-called 'Sabians' of Harran, though not Islamic or Christian or Jewish, were tolerated by the Caliphate as a 'people of the Book'. Many served in Baghdad as mathematicians and astronomers. One of the most famous of their number was Thabit bin Qurra, who wrote a number of extant treatises on various scientific subjects, as well as translating a number of Greek works into Arabic, including the *Corpus Hermeticum*, which served as the Sabians' guarantor of tolerance by being officially recognized as their sacred Book. Evidence has recently emerged that people like Simplicius took refuge at Harran following the closing of the last Platonic Academy at Athens (in 529AD).

This Greek knowledge passed back into the 'West' via Madrid and Baghdad (and a few other cities) following the time of the crusades. Now infused with Arabic algebra, alchemy and astronomy, along with the Zero (and further Indian and Chinese interfusions across the silk road), it was eagerly picked up by many of the leading thinkers of the European Renaissance – up to one of the last of them, Isaac Newton.

With the further development of modern mechanistic science, basing itself in no small part on Newton's own work, this graeco-egyptian 'hermetic' source and origin of science was again repressed -- by an ongoing review process, which shunted it back into the grey.

Interestingly, a century after Newton the German thinker and poet Goethe voiced a criticism of him, directed in particular at Newton's "Optics" and the famous prism experiment, claiming they gave the false impression, by means of an artifice, that light was not one. From the perspective of Newton as alchemist, and translator of the Emerald Tablet of Hermes Trismegistus, the case could be made that

what Newton himself intended to demonstrate was that, indeed, light was one – with one source, whatever colour one might see. Given Goethe's own interest in things alchemical and hermetic, we may assume that he was directing his criticism less at Newton himself and more at the Newton who had passed through a century of 'review process', through the prism of 'review processing', which had, in the course of a century, produced a fractured image of the man, with the hermetic and theological aspects repressed and hidden in the grey.

In order to redress this fracturing, and outline a more complete image of Newton the scientist, I return briefly to his lengthy two part letter to John Locke, the 'Two Notable Corruptions in Scripture' essay, and end with Newton's version of the "Emerald Tablet of Hermes Trismegistus".

The two corruptions in question can be found in I Timothy 3:16 and I John 5:8.

Newton claims these two passages, so often cited to support the Trinitarian view, resulted from inaccuracies in Erasmus' Greek text of 1516. Newton was an ardent Unitarian – ie he believed God was one, not three.

This same 'unitarian' spirit can also be found in the short text of the so-called Emerald Tablet itself. The brief passage is of uncertain provenance, with some arguing for an Arabic origin, some a Greek, and some an Egyptian. Hermetic literature ascribes it to the semi-legendary Hermes Trismegistus himself, who is said to have inscribed it on a tablet of emerald and hidden it beneath the twin pillars of Hermes in Egypt. Other such tablets are also mentioned in the literature – an Hermetic dialogue, rediscovered at Nag Hammadi and presumed to be from the Corpus Hermeticum, called the Discourse on the Eighth and Ninth, ends with the admonition of Hermes to his son to inscribe it on a tablet of turquoise and hide it similarly beneath pillars. Thus, like the 'terma' or treasure texts of Tibetan Buddhists, hidden and then discovered in sacred caves both actual and imaginary, this text claims to be copied and translated from a sacred text originally secreted in Egypt.

What follows is Newton's own translation, which I offer to give some insight into Newton himself as well as into the inspiration or dream that drove him -- and perhaps still drives science today:

Newton's Translation of the Emerald Tablet

It is true without lying, certain and most true. That which is Below is like that which is Above and that which is Above is like that which is Below to do the miracles of the Only Thing. And as all things have been and arose from One by the mediation of One, so all things have their birth from this One Thing by adaptation. The Sun is its father; the Moon its mother; the Wind hath carried it in its belly; the Earth is its nurse. The father of all perfection in the whole world is here. Its force or power is entire if it be converted into Earth. Separate the Earth from the Fire, the subtle from the gross, sweetly with great industry. It ascends from the Earth to the Heavens and again it descends to the Earth and receives the force of things superior and inferior. By this means you shall have the glory of the whole world and thereby all obscurity shall fly from you. Its force is above all force, for it vanquishes every subtle thing and penetrates every solid thing. So was the world created. From this are and do come admirable adaptations, whereof the process is here in this. Hence am I called Hermes Trismegistus, having the three parts of the philosophy of the whole world. That which I have said of the operation of the Sun is accomplished and ended.

From: http://www.alchemylab.com/isaac_newton.htm

Author Information

Dr. Harold von Hofe
CYBEK
32 Medley Lane
Branford, Connecticut 06405
USA

email: h169@sbcglobal.net

Accessing Grey Literature in Public Health: New York Academy of Medicine's Grey Literature Report

**Lea Myohanen, Elizabeth Taylor, Latrina Keith
New York Academy of Medicine - NYAM**

Introduction

Grey literature plays an important role in the publications world, particularly in science and in health care (Cordes 2004, 49-69); (Gray 1998, 842-52). Its use has been studied fairly extensively as evidenced by the number of articles published in library science literature as well as the literature in public affairs. A recent search through Library Literature (DIALOG file 438) returned 90 records about grey literature, 26 since 1999. A similar search in PAIS International yielded 23 records. Since 1993 there have been five international conferences devoted to grey literature.

The Fourth International Conference on Grey Literature (GL '99) in Washington, DC, in October 1999 defined grey literature as "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." (<http://www.nyam.org/library/greylit/whatis.shtml>). Because of its elusive nature, researchers and others interested in grey literature often find it very difficult to locate unless they are already familiar with organizations routinely publishing it. In the United States the Kaiser Family Foundation, the Commonwealth Fund and the Urban Institute are examples of organizations that are regular producers of grey literature devoted to the issues of health care and public health. But outside of the well-established producers the question continues to be "How do I find material that is not indexed by the traditional indexing methods of databases or union catalogs?" The impetus for the start of The New York Academy of Medicine's Grey Literature Report as a finding aid to grey literature came from a conference held at the Academy in 1998 where participants identified grey literature as an integral part of their research resources. Dr. Brad Gray of the Academy's Division of Health and Science Policy subsequently wrote about the importance of grey literature in health policy research (Gray 1998, 842-52). The Academy's Grey Literature Report was first published in 1999, and it originally began as a way to serve the needs of the Academy's internal researchers. Over the years, however, the popularity of the Report has led to the establishment of a subscriber list of over 350 readers who receive the Report via e-mail. The Report is currently published quarterly in an electronic format.

Methods and materials

Collection development in grey literature

The subject areas of The Grey Literature Report span the broad areas of public health, health and science policy, health services research, and health of minorities and special populations, including children, women, and the elderly. As mentioned earlier the foremost audience of the Report and the target of the collection efforts are the researchers at the New York Academy of Medicine. Therefore, we especially seek material dealing with health disparities, health insurance including Medicare and Medicaid, consolidations of hospitals, gun violence, public-private partnerships in health care, immigrant health, and various forms of infectious diseases affecting particularly urban populations, and since September 2001 disaster related material. However, we also collect material of international nature when the subject matter fits within the Academy's research scope so we regularly collect material from the United Kingdom and Canada, from organizations such as the King's Fund, the Nuffield Trust, and the Fraser Institute, and in some instances from other countries.

To date, material in the purely clinical medicine areas has not been collected though we are considering expanding the Report in this direction, particularly as it pertains to clinical material for the researcher and the practicing public health workforce.

Grey literature items added to the Report include case studies, conference proceedings, discussion papers, fact sheets, issue briefs, government documents, research reports, and white papers. It may surprise the reader that we include government documents in the Report; after all United States government documents are indexed and often accessible via the Government Printing Office website. However, as librarians we have come to the conclusion that this database is often too massive for easy navigation and identification of relevant documents produced by the large number of governmental organizations.

So far no audiovisual material, videos or webcasts are included in the Report although there has been discussion about including these publication formats in the future.

Relevant publications are identified by scanning the publications pages of organizations. This is done by two Public Services librarians who also serve as liaisons for some of the research staff of the Academy and are therefore familiar with the local research interests and activities. In some instances the librarians are notified via e-mail by the organizations themselves about recently published titles. Examples of this are the Kaiser Family Foundation, the Commonwealth Fund, and Mathematica Policy

Research, Inc. though email alert does not always include every new publication so direct scanning of the organization is still required.

The grey literature producing organizations are identified through current awareness sites, such as American Healthline and the various daily reports from the Kaiser Family Foundation, staff referrals, and serendipitous reading. Scanned organizations include government agencies, non-governmental organizations such as the World Bank, the Organization for Economic Cooperation and Development (OECD), and the various organizations of the World Health Organization, university research departments, research centers, and international organizations. (A list of publishers is available on the Grey Literature Report web site at <http://www.nyam.org/library/greylitorgs.shtml>). The two librarians charged with the collection development scan these organizations at least once during the publication cycle. At the time of writing this paper we scanned nearly 230 organizations. The Report includes only English language material and primarily only resources that were published relatively recently in order to maintain the current awareness nature of the Report. When adding new organizations the librarians look for material that was published over the last two years.

In some cases we also receive copies of reports from Academy staff.

Format of the Grey Literature Report

The New York Academy of Medicine's Grey Literature Report is a list of publications organized alphabetically by the issuing organization or, at times, the organization on whose server the publication is housed. All the publications that have an electronic equivalent are linked to the corresponding URL so that by clicking on any given title the reader is immediately taken to the full text of the publication. Because of the unstable nature of the URLs the Library also collects paper copies of nearly all of the items identified in the Report. The librarians order free print copies from the publishers whenever this is feasible, i.e. when ordering can be done electronically and results in the document being received in a reasonable period of time. Often, however, the turnaround time is such that, for the sake of expediency, it is preferable to print copies directly from the web. Currently we do not fully print publications exceeding 80 pages but provide our Technical Services department with at least the title page for cataloging purposes.

Technical Services aspects of the Grey Literature Report

Individual publications are cataloged to Library's online catalog and to OCLC. According to 2003 data over 65 percent of the grey literature items added to OCLC were unique at the time of addition. Currently, two catalogers and two technical services assistants input bibliographical information (title, author, organization, etc.) and assign subject headings to each record. Each record gets a local call number beginning with the letters NQ or OCT (for smaller sized publications). In addition, the electronic link and the full table of contents, when available, are added for each publication. Grey literature items owned by the Academy library can thus be located by searching the Academy's online catalog.

Since the fall of 2001 the New York Academy of Medicine has been under contract with The National Library of Medicine (NLM) to identify, collect, and house grey literature material from organizations selected by NLM and to catalog them into NLM's LocatorPlus system. Since the beginning of the contract the New York Academy of Medicine Library has contributed nearly 700 reports to LocatorPlus as well as many serial-like titles to PubMed.

Publication and dissemination

The final task in the publication process is the actual production of the Grey Literature Report. Bibliographic data, including the web address of each document and relevant html tags, are entered into Microsoft Notepad and finally into the Academy's content management system, which produces the web version of the Report. This data entry is currently shared by the collection development librarians and a member of our Technical Services department. For documents which are not available in full text on the web we indicate that these must be ordered from the publisher. Prior to publication all web addresses of the documents are checked to make sure they are functioning. The index page contains a disclaimer stating that all URLs were "live" at the time of publication.

The Report is published on the Report's webpage, <http://www.nyam.org/library/grey.shtml>. It is also easily accessible through the alias <http://www.greylit.org>. The Report is published during the last week of February, May, August and November. All previous editions are available on the website but no maintenance or link checking is carried out on the pages. At the moment we do not have the man power to re-check dead links nor, more importantly, to locate correct URLs and replace dead links with new ones.

The Report is disseminated in 3 ways – on the Grey Literature Report webpage, to the MedLib-L, a medical librarian listserv with over 2000 subscribers worldwide, and via email to the list of subscribers described below.

We do not know how many people access the Report or the list of organizations directly but the site received 291 hits from 231 unique visitors during August of 2004.

Subscribers

In order to facilitate direct access to the Report, a free subscription can be obtained by sending e-mail to one of the public services librarians working on the Report. A list of the subscribers is kept in the Academy's Groupwise account and is updated as necessary.

As of June 2004 there were 385 current subscribers to the Grey Literature Report. Of these about 27 percent come from outside of the United States and 19 percent come from educational institutions. From January 2004 to the beginning of June 2004 we added 67 new subscribers, which is 17 percent of the total. We seem to receive a flurry of new subscribers after the latest Report has been sent to the current subscribers. We also suspect that many of these come from the listserv. This may be an important factor to take into consideration in the future if the Academy Library begins to market the Report.

Subscriber survey

In the fall of 2003 the grey literature team conducted a survey of the then nearly 300 subscribers to the Grey Literature Report. The survey consisted of 11 questions mailed to the subscribers electronically. The response rate was about 28 percent. One goal was to find out whether the subscribers were librarians or researchers involved in public health and health policy. As we suspected, most of the respondents were librarians. In addition, we were interested in finding out how the Report is used. Most of the respondents indicated that they pass the Report along to researchers in their organizations. In terms of improving the Report the survey confirmed much of what we already suspected. It indicated that most subscribers would prefer to receive the Report monthly as opposed to quarterly. Another suggested improvement, that we are aware of and have discussed, is making the Report searchable directly as opposed to searching via the library's online catalog.

Problems and difficulties

One of the most frustrating experiences for any user of the internet is the changeable nature of the web addresses assigned to websites or documents. You can find a document one day but the next day the address you used may no longer be valid. You may be able to locate the document again by using one of the search engines or by going to the web site of the issuing organization. In a worst case scenario, however, the issuing organization has removed the document completely so it no longer is available electronically via the internet. This was confirmed when we visited the site of one organization we regularly scan for documents. Documents are removed apparently randomly to make more room for new ones.

Another example of url changes can happen during the redesigning of an organization's web site. As mentioned before we verify the web address of each document prior to the publication of the Grey Literature Report. Not too long ago one organization that we routinely scan changed the look of its website and in the process changed the addresses of all the documents we had included in the upcoming Report. We didn't become aware of the changes until the last moment and did not have time to locate the new URLs for the documents. We had to remove all the old web addresses and insert a notice to this effect in the Report.

The unstable nature of the urls has led the New York Academy Medicine library to collect hard copies of the grey literature documents since the start of the Report. This way even if the documents are no longer available via the web the user can have access to a print copy through our library. There are, however, several drawbacks to this approach.

- Most libraries today are faced with space issues which make storing paper copies an unattractive alternative.
- There are costs involved for paper and toner used for printing.
- We consider that some documents are too long to print out fully - we limit printing to documents that are less than 80 pages long - and these documents will therefore continue to be available only electronically.
- Access to a paper copy does not necessarily meet the user's need for immediate access to information. The user has to either get in touch with our library or the publisher for a copy and some organizations charge for a print copy even though the electronic version may be available for free.
- We do not have an acquisitions budget for grey literature documents and therefore can purchase only a few print items, which are chosen for their value in supporting in-house research.
- Some issuing organizations do not have a very efficient hard copy distribution system for print copies, which may make getting a print copy frustrating and time consuming.

In addition to these major problems there are many problems which only reinforce the conviction that grey literature is elusive and ephemeral.

Document revisions are occasionally noticed where the revised document has the same URL as a previous version, which is no longer available, or we cannot find, electronically. Therefore, the print copy which we now hold in the library is not the latest document, is not identical to the revised one and appears not to be available. Even were we to notice such a change would we just simply print out the

newer version and replace the old one, would we link the new one to the bibliographic record with a note and keep both versions? Does the old version matter if the author intends it to be replaced with new version?

Document types can be confusing. Some completed reports are called "working papers" but we have also found documents that are labeled "draft" and/or "working paper" and imply a work still in progress. Some clearly state that it is for limited distribution, not to be distributed, quoted or used without permission of the authors. Does this mean we are not to print a copy, and that we are not to include it in the Report? If we do put it in the Report, we are clearly distributing it to our subscriber list. Do we need permission to include such documents and print a copy for the library?

Access problems of a different kind occur with those organizations publishing grey literature that only allow subscribing members access to their reports. They seem to be a minority as most organizations seem devoted to making their research available to all, but there are a few who restrict access. Some national membership organizations fall into this category, e. g. American Benefits Council. Abstracts of documents may or may not be available, depending on the organization. From the information available documents at these association membership sites do not always appear to contain member sensitive or member specific information. Occasionally, documents that at one time were available only to the membership subsequently become freely available. These are issues which originate with the publishing body and which we may not be able to solve.

In-house issues that we have not addressed in great detail include the development of more detailed policies governing the Report. Collection development guidelines are fairly broad and there are currently no retention or deselection policies other than those issued by NLM for the documents that we hold for them. We need to start to examine our collection, and track usage, if possible, in order to define what our long-term retention policies should be.

Future Development

The future areas of exploration for the Grey Literature team include developing a searchable database for the end-user, digital archiving issues and marketing the Report to a wider audience, including public health professionals, academics, and politicians interested in the policy areas of health care.

Searchable database

Currently there are two ways for users to locate grey literature material at the New York Academy of Medicine library. The Academy's library catalog can be searched and quick limits can be applied to restrict the search to grey literature items only. The other is simply for the user to take advantage of the browser's "Find" command while scanning each Report. The inability of our users to search the contents of the Grey Literature Reports in a more robust way has led us to look for other forms of content management. One possibility is to migrate the Report to a stand-alone database which could be searched in a variety of ways, including title, author, subject, publishing organization, keyword, and the date of publication. This necessitates hiring a programmer to develop a database to our specifications. Another possibility is to use one of the freely available databases such as Greenstone.

Solving the problem of lack of persistent URLs

As we saw earlier, there are difficulties with the lack of persistent URLs. One way of getting around the problem of changing web addresses would be to archive documents permanently on a server. This would require that publishers agree to house their documents on a server outside their own organization. This solution, however, entails obvious copyright problems. Who now owns the document? Also, what happens when an update is issued to an already existing document?

Another solution would be to take advantage of the DOI or digital object identifier numbers, which are presently attached to, for example, journal articles available on the web. Because the DOI number is attached to the document itself not to any particular web address, the number can lead the user to the document's current web location (<http://www.doi.org/index.html>). This would require that publishers of grey literature material register their documents for a DOI number. At present, however, only major commercial publishers, such as Elsevier and McGraw-Hill are participating in the DOI effort.

Working with state public health departments

The 2003 survey of the subscribers to the Academy's Grey Literature Report indicated that most of them were librarians who then passed the interesting documents along to their patrons. In the future we are hoping to directly recruit public health professionals as subscribers, particularly those in the health departments in the New York City area, and so increase the readership of the Report. The library's administration has already made preliminary contact with certain interested parties to begin a collaborative effort between the NYAM librarians and the public health workforce. This would benefit public health practitioners by putting them directly in touch with relevant grey literature. An additional benefit of this contact would be in the area of collection development. By sending their recent publications to the New York Academy of Medicine library public health departments could publicize their recent output in the Grey Literature Report and have a repository for their documents. The obvious

advantage to the New York Academy of Medicine library would come from having documents sent to us instead of having to go after them, although we would still like to have the final decision as to whether we will retain a document or not. At the moment we scan the publication pages of the state departments of public health within the Middle Atlantic Region of NLM's Regional Medical Library system.

Grey literature in clinical medicine areas

At the moment we do not collect material in purely clinical medicine for the Grey Literature Report unless there is a public health aspect to the document, such as AIDS, tuberculosis, smoking etc. This is another possible area of expansion for the grey literature effort at the New York Academy of Medicine. The preferred avenue of publication in clinical medicine is journal articles or books. However, grey literature documents exist in clinical medicine as well, and examples of these are position papers, and treatment and practice guidelines put out by organizations such as Alzheimer's Association, American Lung Association and Forum for Collaborative HIV Research. We are not sure how much of the material is not found in the journal literature, or in online resources such as guidelines.gov.

Webcasts, videos, audiovisual material and power point presentations

Currently the New York Academy of Medicine only collects grey literature material in paper format. Yet many of the organizations that we scan for literature also produce material in the form of webcasts, videos, power point presentations and audiorecordings. Many conference proceedings are also captured using these formats in addition to the traditional paper proceedings.

Conclusions

The New York Academy of Medicine's Grey Literature Report has been published in an electronic format quarterly for the past five years. This period has seen an increase in the number of organizations from which reports are culled as well as an increase in the number of subscribers to the Report. Even though an increasing number of grey literature producers publish their reports on the web and hence potentially increase the public's exposure to their research, libraries and users of grey literature are faced with the issue of bad links and reports disappearing from the web. In order to remedy this situation the library at the New York Academy of Medicine collects hard copies of most of the grey literature publications listed in the Grey Literature Report. Future directions for the library include investigation into development of a searchable database and possible archiving agreements with prolific grey literature publishers.

The Report can be read at <http://www.nyam.org/library/greylit.shtml>. Subscriber information is available at this address.

References

- Cordes, Ruth.** 2004. Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an international Marine Scientific Advisory Body. *Canadian Journal of Information and Library Science* 28, no. 1:49-69.
- Gray, Bradford H.** 1998. Sources used in health policy research and implications for information retrieval systems. *Journal of Urban Health* 75, no. 4:842-52.

Author Information

Ms. Lea Myohanen
Reference Librarian
New York Academy of Medicine
1216 Fifth Ave.
New York, NY 10029
USA

lmyohanen@nyam.org
<http://www.nyam.org/library/grey.shtml>

Grey Literature in Special Libraries: Access and Use

Sara L. Ranger
University of Houston

Abstract

Patrons of special libraries are considered heavy users of grey literature. However, little is known about the actual use and users of grey literature in these libraries. This paper explores who grey literature users are in special libraries and why they use grey literature. Also discussed are the specific types of grey literature most often used and how these items and collections are organized and accessed. In this study, only items that librarians physically collected were included in the definition of grey literature; online resources were excluded.

Fifteen librarians from fourteen organizations were interviewed over the course of approximately eighteen months. Four librarians worked at government libraries, four at corporate libraries and five at specialized academic libraries. The interviews concluded with physical tours of each library, so that librarians could present grey literature *in situ*. Tours also allowed the investigator to see how grey literature was physically organized and to identify items that may have been overlooked in the interview.

The amount of grey literature found in each library and in each type of library (academic, corporate or government) was analyzed. Also determined was the size of the grey literature user population in each library and each type of library. The different kinds of grey literature found in each type of library were compared by size and frequency of use.

Among the libraries studied, it was found that grey literature comprises anywhere from less than one to over 95% of total collections. There was less grey literature in the corporate than in the academic and government libraries. Academic libraries reported the widest range of percentages of grey literature in their collections.

Reports, conference proceedings and newsletters were the types of grey literature most often found in special libraries and were also the kinds used most often. Corporate librarians reported that their patrons use conference proceedings more than reports or newsletters, but academic and government librarians reported the opposite.

The special libraries studied made their grey literature accessible through paper and electronic catalogs. Over half of the electronic catalogs were accessible over the internet. While the academic and government libraries studied made their grey literature available to all, either in the library or by interlibrary loan, the corporate libraries made grey literature available to employees only. A wide range of description occurred: librarians cataloged grey literature at the item level or the series level but some items were not cataloged.

Half of the special libraries reported less than forty percent of patrons used grey literature. Of the grey literature users, many used it to generate their own grey literature. They also used it to create commercial and personal writing.

In conclusion, the reasons grey literature is not used by more patrons are discussed; possible factors include the lack of complete description and catalog access, and lack of visibility. It is recommended that special libraries contribute to grey literature depositories, create more cataloging records, make those that exist more complete and more accessible, and cooperate with other special libraries to accomplish these goals.

Introduction

Grey literature has been studied and discussed since the 1970s but users are rarely mentioned. This paper summarizes knowledge gained from interviewing librarians who deal with both the grey literature in their collections and the patrons who use it. Possible improvements and barriers to making grey literature easier for people to find and use will be discussed. Also proposed is the possibility that a lack of widespread and consistent cataloging negatively impacts grey literature use.

Grey Literature Defined

People have defined grey literature in many different ways. The simplest and most consistently used definition can be found in the Third International Conference on Grey Literature, and is now called the Luxembourg Convention on Grey Literature (Farace, 1998, iii). 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." Electronic forms of grey literature have greatly increased the number of grey literature documents. Including electronic items expands earlier definitions. Rather than only physical objects such as paper, video, and microfiche, grey literature now includes a much larger array of documents, including websites, electronic datasets, and electronically published articles and reports. Traditionally, grey literature consists of categories of documents like technical reports, theses and dissertations, newsletters and datasets, often uncataloged and hidden from the public view. Dominic Farace, founder of GreyNet, stated that some documents now available electronically should not be considered grey literature. "Grey Literature must be distinguished and distinguishable from junk e-mail, vanity press, and other gigabytes of e-trash on the net." (Gelfand, 2000, 74) While some electronic documents truly are grey literature, for the purposes of this paper, the Luxembourg Convention definition will be used and electronic documents will be excluded.

Grey Literature Users

Before describing his system for organizing and retrieving grey literature, Jeffery (2000) pinpoints his potential customers or users as researchers, university officials, research council officials, commerce and industry, intermediaries, media and the general public. In other words, in his eyes anyone could be a grey literature user. Auger's (1994) list of organizations that contributed to the now-defunct Committee on Scientific and Technical Information (COSATI), also identifies nearly everyone as a potential user, from government officials, academics and scientists to business people, teachers and people who belong to societies and organizations (3).

Weintraub (2000b) says that grey literature is ubiquitous: citizens need grey literature to make informed decisions about government and the other institutions they deal with on a daily basis. For instance, if their city is planning a road project they might want to look at any related reports commissioned by the city. The literature shows that while researchers are often mentioned as frequent users of grey literature, civil servants, teachers, students and the general public also use grey literature resources.

There are some groups that are heavier users. Fazio (1996) identifies grey literature users at the Central Library of the National Research Council in Rome as researchers interested in the newest information in their fields, teachers looking for pre-digested information already packaged and easy to present, and managerial staff looking for data to support decision-making. Lux (1996) reports that the Senatsbibliothek Berlin (SEBI), a collection that is approximately 70% grey literature, serves German civil servants, researchers, and students studying city and regional planning.

Organizations producing grey literature do not necessarily realize the extent of the audience reading and using their documents. Artus (1994) queried researchers in the social sciences about where they sent copies of grey literature written by themselves and found that over a quarter of the copies went to colleagues with whom the researchers were personally acquainted. If organizations produce more copies of their documents, they will almost certainly be read by a larger and wider audience, but more importantly, by an audience not necessarily anticipated by the organization. Because colleagues share interesting papers with one another, a single grey document can make its way from its intended audience to an audience in the same area but outside the organization, and finally to an audience in related fields.

Why do people use grey literature?

In Brown's research (1999) on astronomers, chemists, mathematicians and physicists, she discovered that while groups of professors in these disciplines do not often use grey literature (pre-prints and conference proceedings) for teaching, they do use it extensively in their research. A survey by the Ministère de l'Enseignement supérieur (reported in Comberousse, 1996), showed that grey literature was important for monitoring changes in technology, for current awareness and as a way to find experts in a field. Bichteler (1991) details the types of grey literature geologists produce and use, including field trip guidebooks, state survey open-file reports ("quick and dirty" versions of more edited reports), research newsletters, maps, and dissertations and theses. She stresses the importance of grey literature resources in geology, where researchers depend on a wide variety of sources to obtain the information they require. In an earlier paper, she and Ward (1989) describe the significance of newsletters, showing how important they are within specialist groups such as paleontology organizations, where findings can be reported long before they appear in formal journal articles.

Method

Interviews

In order to discover the current state of grey literature in specialized libraries, special librarians were asked how much grey literature their libraries hold, who uses it and for what, and how it is treated in terms of providing access. Interviews about grey literature were conducted with librarians between

February 2002 and May 2003 in Washington, Michigan and Texas. The interviews began with a set of seven questions (see Appendix I) but were often followed up with further questions and librarians were encouraged to let their responses lead to areas not defined by the set questions. Answers to each question were written down and the interviews were taped. Interviews generally lasted between fifteen and thirty minutes.

Sixteen librarians from fourteen organizations were interviewed. One interview was not included in the results due to a complete lack of paper grey literature in the library. Of the thirteen interviews used, four were conducted at government libraries, four at corporate libraries and five at specialized academic libraries. In each case, the interview was conducted in the librarian's office or in the library itself and was followed by a tour of the library.

Librarians and libraries were identified as subjects in one of three ways. Several were librarians already familiar with the author. Several were referred by friends, family and colleagues. Two volunteered to participate after the project was presented at a professional meeting.

Results

The majority of the libraries in this study have substantial collections of grey literature that make up at least ten percent of their total holdings; holdings ranged from <10% to over 90% (see Table 1). Four libraries reported that less than 10% of their library holdings consisted of grey literature. Academic libraries have the greatest range of grey literature holdings, and corporate libraries have a smaller range and less grey literature overall (see Table 2). Twelve of the thirteen estimates were made without reference to the collection. In other words, the librarians were simply estimating using their knowledge of their libraries. The thirteenth estimate was based on a recent inventory and could be considered the most accurate. The librarians reported a similarly wide range of estimates of the numbers of patrons who use grey literature. They estimated that less than 1% up to 90% of their users takes advantage of their grey literature collections. Both academic and government libraries estimate a wide range of users accessing grey literature, while corporate libraries report a smaller range and fewer users in general.

Table 1. Grey Literature Collections and their Use in Libraries

	Number of Libraries Reporting				
	<1- <10%	10- 20%	20- 40%	>40%	Total
% total collection reported as being grey literature	4	4	3	2	13
% of total user population reported as using grey literature	4	1	3	3	11*

*Two librarians responded "not sure" to this question

Table 2. Grey Literature and its Use in Three Types of Libraries

	Type of Library		
	Academic	Governmental	Corporate
% total collection reported as being grey literature (range)	<1-90%	10-50%	<1-30%
% total user population reported as using grey literature (range)	<1-95%	1-90%	5-50%

Librarians most often reported owning three types of grey literature: reports (of various kinds); conference proceedings; and newsletters (see Table 3.). Both reports and newsletters are available in a majority of the libraries visited, but reports are not available in as many corporate libraries as government and academic libraries. Government libraries tend to collect theses and dissertations as well as maps and atlases, but these two types of resources were not available (or possibly were not considered to be grey literature) in other libraries. While academic librarians did not mention ephemera, one government and one corporate librarian mentioned having ephemera in their collections.

Overall, reports were overwhelmingly cited as the most often used type of grey literature (see Table 4). Newsletters and conference proceedings were mentioned half as many times. One librarian stated that no one type of grey literature is used most often in his library, but that resources of any sort that cover a few specific issues and topics tend to be used frequently. Several librarians named report series as highly used grey literature. These were always series covering a specific topic, such as the International Whaling Commission reports. Corporate library patrons seem to use conference proceedings more often and newsletters less often than government and academic library patrons.

Table 3. Types of grey literature available in special libraries

Type of Grey Literature	Total No. Libraries reporting	Academic Libraries only	Government Libraries only	Corporate Libraries only
Reports	7	3	3	1
Conference proceedings	6	1	1	4
Newsletters	6	2	2	2
Theses & Dissertations	3	0	3	0
Atlases & maps	3	0	3	0
Ephemera (pamphlets, fact sheets, letters)	2	0	1	1

Table 4. Types of grey literature used most often

Type of Grey Literature	Total No. Libraries reporting	Academic Libraries only	Government Libraries only	Corporate Libraries only
Reports (government)	6	3	2	1
Newsletters	3	2	1	0
Conference Proceedings	2	0	0	2
Topical Resources	1	0	1	0

The holdings of all but one of the libraries studied are at least partially accessible using an online catalog. Of those, over half are accessible to anyone with access to the internet. Three libraries' holdings are accessed using both online and paper catalogs (card or shelf list) and one is accessible by paper catalog only. Within all types of cataloging systems, varying numbers of grey documents are cataloged. The largest collection of uncataloged individual grey literature items is probably the microfiche collection of government reports at one academic library. Some of the librarians catalog each and every item in their collections, although they may catalog items at the series level rather than at the item level. Other items or even series may be completely uncataloged. When asked what part of the grey literature collection was used most, one government librarian mentioned a specific report series, which also happens to be uncataloged. Its users simply have to know it exists or be guided to it by the librarian. If they know to ask for it, the librarian can show them where it is and help them locate documents within the series, but for the most part the users must have prior knowledge.

Academic and government libraries make their grey literature collections available to the public, either in the library or by interlibrary loan, but corporate library collections are available to employees only (or to other corporate librarians via interlibrary loan).

One thing that arose during the interviews was the fact that most of the users of grey literature also produce it. In fact, they often use it in order to make more. However, being the author of a particular document does not necessarily mean that the document can be located once outside the control of the author. Even grey literature producers need help finding grey literature.

Librarians reported that their patrons use grey literature in a wide variety of activities. They use it in their research, to write (often more grey literature), because they have an interest in a topic, for class assignments, as a record of past practices and what an environment used to be like, for localized studies, and for creating models and policies.

Discussion

Why Isn't Grey Literature Used More: Cataloging

Even in the thirteen libraries visited, a large amount of grey literature remains uncataloged, which makes finding and using it difficult. It was clear from the interviews that cataloging is done at a variety of levels, from no cataloging, to cataloging reports at the series level, to cataloging each document separately.

COSATI, the Committee on Scientific and Technical Information (now defunct) created a scheme for cataloging reports. Each record included elements essential for the document's retrieval. There were twelve elements, including typical ones like accession number and title, as well as more specialized fields to describe things like availability and security classification (Auger, 1998). Furthermore, ISO 10444 is now available to guide the numbering of reports.

Some of the librarians interviewed said that they did not do retrospective cataloging because they did not have the funds. While this is an understandable situation and normal budgeting may not allow for a retrospective cataloging project, there may be special funds available, either within or outside the organization the library serves. Furthermore, Farace (1997) reminds us that we are adding to the potential value of grey literature when we catalog it, whether or not it is later accessed and used.

The other major reason that grey literature goes uncataloged is a lack of skilled employees with time to do the work. It is true that cataloging literature created outside of the main stream of publishing is difficult. No cataloging in print (CIP) data is available and it is less likely that cataloging records are available for export. If records are available, they may be hard to find due to a lack of grey literature cataloging standardization. Therefore, cataloging grey literature often translates to original cataloging, which requires more time than copy cataloging and a trained cataloger rather than a skilled copy cataloger.

Report Literature

Most of the librarians interviewed mentioned report series as being important, both because they make up a large part of their collections and because those series are well-used. In a 1992 study of grey literature citations in library and information science journals, Alberani & DeCastro Pietrangeli (1995) found that reports were the most cited type of grey literature. Sidwell, et al. (2000) postulate that reports are important because they contain specific information and also because they contain experimental detail that is not included in other publications, including formal journal articles about the same research. Research by Blagden, et al. (1994) showed that 40% of surveyed engineers rated reports as a very important resource, with only books rating higher. Because of this research, Sidwell, et al. (2000) chastise libraries for "treat[ing] technical reports series as periodicals and hav[ing] only a single catalogue entry at series level..." (131)

Why Isn't Grey Literature Used More: Visibility

Examples from the literature show that the potential user base for grey literature is limitless. Anyone around the world who has an interest in anything governments are doing, anyone who works for a company that produces internal reports, and anyone doing research or just reading a newspaper will run across mention of an obscure document, often referenced only by a number. Why, then do more people not take advantage of the masses of grey literature produced each year? As Schmidmaier (1986) notes, "For the user, the only available information sources are those he can 'see'." (107) In other words, grey literature must be shown to its potential users and it must be readily available. Auger (1998), discussing the low rate of request for grey literature at the British Library, proposes that, "Whilst a lack of awareness of the value of grey literature is without question one factor, obsolescence or perceived obsolescence could well be another." (47)

When these documents are referenced by someone outside the organization that produced them, potential users often have a difficult or impossible time tracking them down because they are not referenced in a descriptive manner. This could occur because the writer did not intend her work to be viewed by anyone outside a specific audience, all of whom would understand the reference. Some writers also reference literature from literature reviews and do not ever see the original material, which can lead to incorrect citations being perpetuated.

Who can actually put grey literature to good use, then, if it is difficult and time consuming to find? At the very least, the intended audience should be able to find and use grey literature. For instance, an environmental consulting firm creates a report for the commissioning organization. Most likely that organization will read, file and perhaps use the report to act or to create further documentation. Research reported in an organizational newsletter will reach its intended target: the organization where the research was conducted. The newsletter itself may be filed, recycled or forwarded to colleagues in other organizations. Thus, grey literature is consumed by people in a small arena, but it can easily find its way further afield, and this is where its use, collection, organization and retrieval become a challenge. Meier & Greenberg (1994) demonstrate how difficult it is to find unpublished information produced by a utility. They had to go through individual filing cabinets to locate the data upon which manufacturers based their product claims. The information existed, but it was nearly impossible to locate.

An obvious solution would be for producers of grey literature to make sure their literature does not leave the organization. If the information is truly internal and meant to remain internal, for example in companies attempting to maintain a competitive advantage and government organizations working with sensitive materials, this might be feasible as well as desirable. However, most organizations either do not consider or do not care when their materials find their way outside the intended realm of users. Some might be happy to know that their documents are useful to others. Even if the organization does care, it is difficult to control the spread of information.

Another solution, of course, would be to provide a way for organizations to send copies of their grey literature to a central repository. The repository would take care of organizing and describing the documents. Its information specialists would also help potential users find documents and send the appropriate materials to users or make them available electronically. In fact, there are many such centers around the world, including the SIGLE database (System for Information on Grey Literature in Europe) produced by EAGLE (European Association for Grey Literature Exploitation), the British Library

Document Supply Centre, GRISELI (standing for French words for grey and literature) centers in France, and NTIS (the United States' National Technical Information Service). Have these repositories increased the number of people able to use grey literature? Do users know about these centers or do they still use local resources such as public, academic and special libraries?

It appeared from the interviews that the longer a librarian had been with his or her library, the more confident he or she was in answering the interview questions. Obviously, the longer a librarians works in a library, the more familiar he or she becomes with its collections. This familiarity also translates into expediency in suggesting and finding the grey literature within the collection. The more familiar a librarian is with her collection the more able she is to share its parts with users. When there is no librarian, who can help a user find grey literature? Meier and Greenberg (1994) note that in the case of grey literature produced by utilities, "The only way to find the gray literature was to ask the staff."

One partial solution is to increase the availability of reports through the internet, especially when they can be downloaded in their entirety. Gelfand (1998) notes that the internet has already allowed for better visibility. However, this does not solve the problem of users not knowing that these reports exist. Grey literature in general must be publicized by information professionals (Aina, 1996). They are the ones controlling, in large part, what their users find, so making grey literature a priority within a library would certainly improve its visibility.

Conclusions

Depositories

If each country or region chooses a depository for its grey literature, publicizes the depository, the depository actively solicits documents from as many sources as possible and describes and catalogs its collection, grey literature as a whole would be in a good position to become a well-used resource. Because grey literature is important around the world, an international standard for cataloging, such as the one created by COSATI might be vital.

Cooperation

Because the author volunteered at one of the government libraries, she was already aware that the librarians there make use of the local university's catalog and collection when one of their users needs an item they do not have. Schmidmaier, a German information specialist, said that "national libraries cannot make decisions about 'grey literature' without cooperation with special libraries." (104) The opposite may also be true, in large part because the special libraries are the first and sometimes only repositories for grey literature produced by the organizations they support. Maintaining free or inexpensive loan policies like those the government libraries use is also important.

Recognition

If depositories are established (or maintained where they already exist), users have to know about them in order to take advantage of their resources. Organizations that collect grey literature also have to make sure they recognize their own users. Depositories should contact universities, public libraries, and make sure that their contributors are advertising the depository's existence to their organizations. They should strive to have a presence in the communities they serve, so that grey literature can gain the recognition and use it deserves.

Acknowledgments

I would like to thank Catherine Stickley, Maureen Woods, Brian Voss, Sonja Kromann, Craig Wilson, Louise Richards, Mel DeSart, Sara Davis, Cathryn Bowie, Diana Houston, Molly White, Pat Parker, Anthony McMillan, Daran Bishop, Lee Hilyer and Marilyn Goff for participating in my interviews and showing off their lovely libraries.

References

- Aina, L.O. Promoting the exploitation of grey literature in Africa. In Second International Conference on Grey Literature. Catholic University of America, Washington, D.C.. November 2-3, 1995. Amsterdam: Grey Literature Network Service. 63-72. 1996.
- Alberani, Vilma and Paola DeCastro Pietrangeli. Grey literature in information science: production, circulation and use. *Inspel*. 29(4). 240-249. 1995.
- Artus, H. Grey literature and the author/respondent relationship. In First International Conference on Grey Literature. RAI Congress Centre, Amsterdam. December 13-15, 1993. Amsterdam: Grey Literature Network Service. 290-307. 1994.
- Auger, P. Grey literature: establishing a clear identity. In In First International Conference on Grey Literature. RAI Congress Centre, Amsterdam. December 13-15, 1993. Amsterdam: Grey Literature Network Service. 1-4. 1994.
- Auger, C.P. *Information Sources in Grey Literature*. London: Bowker Saur. 1998.
- Bichteler, Julie. Geologists and gray literature: access, use, and problems. *Science and Technology Libraries*. 11. 39-50. 1991.
- Bichteler, Julie and Dederick Ward. Information-seeking behavior of geoscientists. *Special Libraries Association*. 80(3). 169-178. 1989.
- Blagden, J., J. Harrington and H. Woodfield. Eurilia (European Initiative in Library and Information in Aerospace): an audit of aerospace information needs in five European countries. Cranfield University: COA report no. 9405, November 1994.
- Brown, Cecelia M. Information seeking behavior of scientists in the electronic information age: astronomers, chemists, mathematicians, and physicists. *Journal of the American Society for Information Science*. 50(10). 929-943. 1999.
- Comberousse, Martine. Grey literature in France: the GRISELI Program. In Second International Conference on Grey Literature. Catholic University of America, Washington, D.C.. November 2-3, 1995. Amsterdam: Grey Literature Network Service. 126-132. 1996.
- Farace, Dominic J. *Notebook on Grey Literature: a Systematic Approach to Grey Literature in Printed and Electronic Form*. Amsterdam: TransAtlantic. 1997.
- Farace, Dominic J. Foreward. In Third International Conference on Grey Literature. Jean Monnet Building, Luxembourg. 13-14 November 1997. Amsterdam: Grey Literature Network Service. iii. 1998.
- Fazio, Annunziata. Organization and development of a national center for coordinating and disseminating scientific, technological and industrial information in Italy. In Second International Conference on Grey Literature. Catholic University of America, Washington, D.C.. November 2-3, 1995. Amsterdam: Grey Literature Network Service. 238-245. 1996.
- Gelfand, Julia. Teaching and exposing grey literature: what the information profession needs to know—examples from the sciences. *Collection Building*. 17(4). 159-166. 1998.
- Gelfand, Julia. Interview with Dominic Farace, founder of GreyNet. *The International Journal on Grey Literature*. 1(2). 73-76. 2000.
- Jeffery, Keith G. An architecture for grey literature in a R&D context. *The International Journal on Grey Literature*. 1(2). 64-72. 2000.
- Lux, Claudia. Local government and consultants: a grey literature database on local reports in the Senatsbibliothek Berlin. In Second International Conference on Grey Literature. Catholic University of America, Washington, D.C.. November 2-3, 1995. Amsterdam: Grey Literature Network Service. 133-145. 1996.
- Meier, Alan and Steve Greenberg. Utility programs: a journey through the gray literature. *Home Energy Magazine Online*. July/August 1994.
<http://homeenergy.org/archive/hem.dis.anl.gov/eehem/94/940709.html>
- Schmidmaier, Dieter. Ask no questions and you'll be told no lies: or how we can remove people's fear of "grey literature". *Libri*. 36(2). 98-112. 1986.
- Sidwell, C.A., P.A.S. Needham and J.D. Harrington. Lightening grey literature: making the invisible visible. *The New Review of Information Networking*. 6. 121-135. 2000.
- Weintraub, Irwin. The role of grey literature in the sciences. *ACCESS*. 10. Not paginated. Fall 2000b. <http://academic.brooklyn.cuny.edu/library/access/greyliter.htm>

Appendix I

Survey:

1. What types of grey literature are available in your library/organization?
2. How much grey literature do you have? What percentage of your library is made up of grey literature?
3. What percentage of users takes advantage of grey literature?
4. What kind of grey literature is used most?
5. Why do your users use grey literature? E.g. education, supporting writing, fact-finding
6. How is your grey literature organized?
 - a. If it is cataloged, are bibliographic records available?
 - b. Are they print-only or can they be accessed online (or both)?
 - c. Are the records available to people outside the library/organization?
7. Are users able to check out grey lit materials?
 - a. If so, is all available for checkout, or is some reserved for library use only?
 - b. Do you make grey literature available for inter-library loan?

Author Information

Ms. Sara L. Ranger, MS
University of Houston
114 University Libraries
Houston TX 77204-2000
USA

sranger@uh.edu
www.uh.edu

Describing Grey Literature Again: A Survey of Collection Policies

Heather Lehman

MLS Student, Information School, University of Washington

Janet Webster

Head Librarian, Guin Library, Oregon State University

Abstract:

At Oregon State University, we are exploring the differences between collections formed by digitizing existing materials on our shelves and those created by identifying and acquiring digital objects beyond our library. One example is the proposed Tsunami Digital Library (TDL) that would support a growing research program with users needing access to diverse information, much of it grey. The TDL is conceptualized as an information portal designed to coordinate the access and distribution of Internet based tsunami related information. The distributed collection is maintained by contributing partner sites within a digital library system. Yet, defining this digital collection presents challenges that differ from our traditional approach to collection development as the audience goes well beyond OSU, materials are created and housed elsewhere, and the formats are not consistent.

Building on our needs assessment of potential users of the TDL, we surveyed librarians active in the field of digital collections on their practical and theoretical approaches to collecting digital grey literature. Their insights and a careful review of the literature of digital grey collections provide a context for creating a specific collection development policy. Our experience is transferable to any librarian collecting digital grey literature.

Planning for a Digital Collection of Grey Literature

Digital collections of Internet resources place important information close at hand. They facilitate the dissemination of new research faster, and create access for wider audiences. Rather than searching the entire Internet, a collection limits resources to those pertinent to a subject. A collection also suggests that material has been identified as useful by a qualified librarian or subject expert.

The emergence of digital collections and the increasing amount of digital grey literature challenges the library's ability to consistently collect information using our traditional tests and tools. Some of the standard criteria of authority and relevance to the institution's mission remain critical. Other standards, such as longevity and access, are more problematic to apply. One challenge is the lack of consensus among librarians on a definition of grey literature. Another is the failure of collection policies to adequately address grey literature, especially in light of the digital environment. Addressing both challenges is essential to move forward with collecting digital grey literature in a coherent fashion within the university.

At the Oregon State University (OSU) Libraries, this is apparent as we develop digital collections that venture beyond digitizing our special collections and archives. The Tsunami Digital Library (TDL) is one example. It is conceptualized as an information portal designed to coordinate the access and distribution of Internet based tsunami related information. The OSU Librarians in collaboration with colleagues from the Department of Electrical Engineering and Computer Science are testing the utility as well as the feasibility of a distributed digital collection maintained by contributing partner sites with OSU acting as the facilitator. The OSU Libraries also will provide the collection content by evaluating existing digital material, potentially converting useful print material into digital format, and negotiating access with partner sites.

The TDL challenges our expertise with collection development as the intended audience is not local, the potential content is not discipline or program specific, and the formats are not consistent. The audience of tsunami researchers and hazard managers is distributed throughout the Pacific Rim in a variety of settings from labs to local emergency services units. The information is interdisciplinary, real-time and historic, data as well as narrative. Much of it is digital, although not all is accessible to the broad tsunami community given format, language, licensing, and poor web interfaces.

To tackle the challenge, we conducted a needs assessment of potential users to assist us in describing the information they wanted and created (Lehman and Webster, 2004). We surveyed and interviewed forty identified experts within the diverse tsunami community. The survey revealed that this community is a sophisticated user of Internet search engines and tools, and is very familiar with existing information. Not surprisingly, a significant portion of the materials required and desired by the interdisciplinary tsunami community is grey, and these are not adequately housed, maintained, or organized by existing partner sites.

The needs assessment helped us identify the core audience, the digital data, and desirable information resources for the basic collection of the TDL. From the results, this community is looking for a centralized means to find and retrieve real-time as well as historical data, geographic resources, outreach

materials, and modelling information. Researchers would benefit from a collection of resources instead of continuing to resort to searching across multiple disciplines and the entire Internet for information. Given the high volume of grey materials, the breadth of subject areas, the multiple languages, and variety of formats, a collection policy statement could help us and our potential partners better develop the TDL content.

Collection Development Policies and Digital Grey Literature

Opinions vary about the utility of collection development policies with some contending that they quickly become inflexible and outdated (Snow, 1996). Most see collection policies as devices that help libraries target resources perceived to be useful to our audiences and our institutions (Intner, 1996; Lee, 2003; Mack, 2003; Spohrer, 2003). To the inexperienced librarian or one taking over a new subject area, collection development policies can be lifelines while gaining familiarity with a discipline and its local audience. Collection policies traditionally address grey literature tangentially. They describe sources (e.g. societies, government), and formats (e.g. reports, proceedings). They may give geographic scope and language limitations that direct a selector away from some sources and towards others. In general, collection policy statement literature focuses on acquisition decisions involving expenditures, neglecting the plethora of free, digital grey literature.

While working on the TDL statement, we looked to see how others addressed the interaction between collection development policies and digital grey literature. Peters (2001) and Edwards (2004), among others, underscore how collection development is shifting under the pressure of electronic resources. Discussion of consortial agreements, shifts from print to digital, and access versus ownership dominates the digital collection policy conversation (White and Crawford, 1997; Kyriallidou, 2000; Digital Library Federation. [2004]). The policy discussion avoids explicitly addressing grey literature, and the digital collection conversation focuses on tools and technological solutions. Many have discussed the difficulty of defining what is grey, especially in a digital context (Di Cesare and Sala, 1998; Gokhale, 1998; Gelfand, 2000). This difficulty appears to have inhibited the development of coherent policies and practices for treating digital grey materials.

There are exceptions to this dearth. These tend to either address organizing digital grey literature with subject guides that link to digital resources selected according to policy criteria (Latham, 2002), or making grey literature in current collections available (Russell, 2004; Drake, 2004). Additionally, there is some collection policy work in the digital library environment. In this realm, statements generally emphasize format, pricing, and collecting all literature necessary to support research in the particular field (ARL, 2004; Digital Library Federation, [2004]). Many academic libraries have only recently started to develop digital policies to reflect access and ownership issues, technological changes, and the development of interdisciplinary subject areas.

Rather than being discouraged by the lack of models for a digital grey literature collection, we were convinced that the process of developing a collection development policy was useful. The process helped us clarify the scope and content of the TDL, and made us discuss the tests of authority, access, timeliness, relevance, and longevity. However, we were stymied by certain issues:

- Accessing the breadth of digital information when the lowest common denominator is that it is digital;
- Describing the responsibilities of ownership;
- Maintaining quality and authority when conferred by users, producers and librarians;
- Selecting the relevant given the volume of digital grey literature.

So, we turned to our colleagues and surveyed 20 librarians we felt would have insight on these issues as well as the challenges of incorporating digital grey literature into collections.

Survey on Grey Literature and Collection Policies

Through the literature and current policy work, we identified 20 librarians actively involved in some form of digital collection development. These 20 librarians were given a 12 question survey crafted to elicit longer responses (Appendix 1). Selected ones received follow-up correspondence to further clarify their practical and theoretical approaches to collecting digital grey literature. Additionally, the digital collection policies for their respective institutions and subject areas were reviewed for treatment of grey materials. While our survey is small and limited to librarians in North America, it does present some new insight about how digital collection policies handle grey literature.

Analyzing the survey results in a systematic way was problematic given the nature of the questions; they did not, for the most part, generate simple yes or no answers. Rather, we looked for patterns in the beliefs and attitudes that librarians have towards their collection policies and grey literature. We also were interested in working definitions of grey literature in the electronic collection policy setting. We anticipated some validation of the importance of digital grey literature, and the need for collaborative efforts to improve access to these materials. The following presents the results in clusters of survey questions that address the following:

- Subject areas
- Definition issues
- Selection issues

- Cooperation issues
- Access issues

Subject Areas – Questions 1 & 6

The sciences are considered to be the primary source for grey literature. Most of the participants are responsible for science collections so, the awareness of grey literature is high. Seventeen respondents considered their subject areas the highest producers of grey literature. Many respondents note that new technologies are promoting the development of grey literature in fields beyond the sciences.

Definition Issues – Questions 2, 3, 4 & 5

In defining grey literature, every librarian mentions the themes of “visibility / accessibility,” “authority / peer reviewed” and “not traditionally published,” though not all use the same terms. The most precise definitions of grey literature are articulated by the three respondents whose digital collection policies incorporated the treatment of grey materials. More general definitions come from the 16 librarians working with policies that avoid explicit treatment of grey materials, digital or otherwise. One librarian quotes the definition adopted at the Third International Conference of Grey Literature (Farace, 1998); however, this respondent’s policies make no mention of grey literature.

The respondents do not consider foreign language materials grey simply by virtue of being non-English, except when their policies dictate such treatment. Most consider much digital material as grey, but difference in shades of grey emerge. For example, one respondent indicates that the transitory and invisible materials on the Web are obviously grey while another perceives that because the Internet provides greater access to materials, it makes them less grey. Questionable lifespan, accessibility, and on-line publishing are mentioned as why digital materials are often grey.

Respondents divide into three groups when discussing the affect of the Internet on the volume of grey literature. These conflicting responses are again tied to definitions of greyness. Librarians who use a definition focused on access perceive a decrease in volume as the Internet improves access. If defined by publishing methods, the volume of grey literature increases because publication through the Internet is non-traditional. The volume is perceived as steady for those using a definition that combines access and publishing method. Though much digital material is grey, participants note that the Internet is not the only vehicle for grey digital materials; obsolete formats such as tape drives and even CDs render information grey as well.

Selection Issues – Questions 7, 8, 10 & 11

Though the answers do not provide any conclusive process for selecting grey materials, they help address the collection of digital materials in general. In only one instance did the policy of respondents specifically address grey materials. The majority of respondents acknowledge that their collection policies do not specifically address grey materials, and that they do not systematically collect grey items. Those librarians that contradict this trend do so because they serve an audience that mandates that grey collections (i.e. agency technical reports collections) or because they are not academic research librarians (i.e. documents staff). Respondents note that they do not deliberately exclude grey materials, and that they collect it if requested by faculty or if the materials are published by their institution. A few respondents note that they monitor on-line sources for pertinent materials, or have developed a means of reviewing the materials retrieved by faculty visiting conferences and meetings.

The majority respond that grey materials are selected according to the same criteria as any other item: relevance to the subject area and demand by the collection audience. Outside of relevance as a criterion, knowing that an item could be catalogued is the second highest factor affecting selection. Knowing whether an item is indexed or not does not affect selection uniformly because some define greyness by whether the item is indexed or not.

Given the rise of grey materials stemming from increasing interdisciplinary research communities (Wilson and Edelman, 1996; Herring, 2002) and the capabilities of the networked research environment, it is surprising that the majority of responses show a limited number of policies in place to address multi-disciplinary areas. From the responses, strategies for selecting interdisciplinary materials seem to be far more prevalent than collection policies.

Collaboration Issues – Question 11

The concept of using the library catalogue to describe materials housed outside the library appears to be of interest, but not widely applied. Several participants mentioned branch library collections or union catalogues as a means of broadening coverage. Two specifically mention non-library collections that are included in the local catalogue. Collaboration seems to be among libraries or within the institution in terms of integrating access to collections.

Access Issues – Question 12

Eighteen respondents feel that people do not have adequate access to relevant grey literature. Even with ready access through the Internet, grey materials are by their nature difficult to find. As one

participant writes, "A researcher, with a good grasp of their discipline, will know more about the grey literature than the librarian. A student will not. An uninformed citizen will not." Access depends on knowing how and where to search. This requires searchable metadata, some subject knowledge, and coordinated information management.

Twelve respondents identify available resources for grey materials that meet the above requirements. These include Agricola, Virtual Technical Report Center, Scientific Network and Information Network (STINET), Citeseer, and Cogprints. A few suggested that libraries need to consider specialized web portals and harvesting services to facilitate access to materials that are not necessarily catalogued or collected.

Observations on the Survey Results

It was surprising that stronger patterns and more cohesive answers did not emerge given the similarity in background of most of the respondents as well as their experience with electronic or digital collections. The most prevalent approach to selecting grey materials focuses on relevance to a subject area within the confines of an informal definition of what constitutes greyness. As the definition varies, selection and collection practices shift. With the increasing availability of digital grey materials the reliance on an un-stated definition seems inadequate to capture the full array of relevant grey materials in any subject area, let alone overlapping disciplines. It also seems insufficient to assume that relevant grey literature will be collected utilizing existing strategies rather than a specific collection policy.

We originally hoped that the survey and review of policies would yield a strong working definition of grey and practical selection guidelines that would validate the relevance of grey materials for inclusion in a collection. What we found is that grey items are not dismissed but are usually given a low priority by selectors. This is due to the difficulty of dealing with them in a concerted fashion as well as budget restraints that direct resources towards mainstream resources. Digital grey literature reminds us that grey literature in general is messy, and takes time and expertise to collect. The low prioritization of grey literature, digital and print, reflects that many libraries still "fear to introduce how information with no price tag on it can have incredibly high value placed upon it" (Gelfand, 2000).

Currently, the most familiar approach to improving access to digital grey materials is linking and not conscious collection development. These linkages and the development of portals indicate awareness that collaboration among institutions provides greater access to more grey information. However, other than acquiring items at faculty request or selecting on an "I know it when I see it" basis, academic research libraries are still primarily relying on other entities to collect grey materials. This reliance on others to collect so that linking works as a basic strategy is dependent on high maintenance and constant review.

The issue of collaborative collecting underscores the importance and utility of a collection policy. In a collaborative project that extends beyond the borders of a single library, the collection policy can define the responsibilities of ownership and not just the means of access to digital grey materials. The collection policy can be developed to address the responsibility of each institution, including how or when ownership and of collected digital materials should be transferred to or shared with another institution. This is especially important to preserve access to still relevant but endangered digital materials and resources. Approaching collection policy development in this manner can benefit institutional repositories as well as any other digital collections. It also remedies the unsatisfactory approach of providing links to relevant resources with poor interfaces by collecting those pages that are the most important.

The audience for digital collections, particularly those used by interdisciplinary researchers, can be much larger than that of traditional print collections. This expanded audience presents the opportunity to utilize their expertise to identify pertinent materials, thus building upon the traditional method in collection development of utilizing faculty requests as a selection aid. For digital grey materials, deliberate collaboration with the audience may stretch the comfort zones of many librarians. Rather than viewing the collaboration as removing the authority and responsibility of the librarian, it should be seen as a mechanism for identifying and capturing relevant materials from the vast and growing amount of items available.

This new approach to evaluating and conferring authority assures the relevance of materials in the collection by utilizing the expertise of both the librarians and the user community. It would aid in capturing a wider array of digital formats than would normally be included for evaluation. It would deliberately include the expertise of interdisciplinary researchers extending the expertise of the selecting librarian. Since there has been an increase in publication and dissemination of materials directly from the producers of grey literature and other materials, establishing the means to work directly with users rather than relying on the traditional means of evaluating and collecting becomes necessary.

Final Thoughts on Collection Policies and Grey Literature

A heightened awareness of the changes stemming from technological innovations is apparent from the literature and the survey responses. This awareness includes the increased availability of digital grey information. However, the transition from awareness to acknowledgment of the validity of collecting grey literature is less obvious. Instead, there is still an emphasis on providing access to grey materials collected by others and a reluctance to expend time and resources on including grey resources in existing collections. It is not that librarians disregard grey literature. The difficulty and missed opportunities lie in the reliance on strategies that treat grey literature inconsistently.

Collecting digital grey literature means a commitment not only to acknowledging the importance of this literature, but to changing existing collecting strategies into policies that advocate its integration into collections. Once a decision to collect has been made, the research on collaboration and varieties of consortial agreements can help guide the mechanics of collecting of all formats of digital materials. What is needed is more focused discussion on collaboration specifically in regards to grey materials. This discussion needs to include how collaborative agreements can be structured with users and producers of grey literature so that issues of quality and authority are addressed as well as issues of access and ownership.

One way to begin that discussion is to turn to the familiar practice of creating a collection development policy. As a tool, the policy is a solid way to discuss and refine the direction and strategies of creating any collection. It can provide a strong framework for acknowledging that grey materials are a vital part of our digital collections, and outline the mechanisms for integrating them.

Appendix I

Gray Literature Survey

1. What subject areas are you responsible for in the library's collection?
2. How do you define grey literature?
3. Do you consider foreign language materials as grey literature?
4. Do you consider a lot of digital material as grey literature?
5. How has the advent of technologies such as the Internet affected the volume of grey literature – has it increased or decreased?
6. Do some subject areas or disciplines generate more grey literature than others?
 - a. If so, which fields do you consider as having more grey literature?
 - b. Where do your subjects fit into this continuum?
7. Does your subject area collection policy address grey literature?
 - a. If yes, why?
 - i. Do you actively collect it?
 - ii. How do you find grey literature relevant to your collections and how do you evaluate that material?
 - b. If no, why not?
8. Do your collection policies cover digital material that might be considered grey literature?
9. How does your institution handle collection development in overlapping disciplines or interdisciplinary fields?
10. When making decisions about whether to collect a grey publication, what factors make a difference to you?
 - a. Knowing that other departments/ areas within your institution have (print) copies
 - b. Knowing that you can get it catalogued
 - c. Knowing that it is indexed
11. Does your catalog describe materials housed in other areas / departments?
12. Do you feel that people have adequate access to the grey literature that they would find relevant to their work/needs?

Author Information

Ms. Heather Lehman
Atrocity Arts
2009 Harney Street
Vancouver, Washington 98660
USA

heatherlehman@comcast.net
<http://www.washington.edu>

References

- Association of Research Libraries. 2004. *ARL Collections & Access Issues*. Available at: <http://db.arl.org/CAsort/>
- Di Cesare, Rosa & Sala, Cesare. 1998. The Use of Grey Literature in the Agricultural Economics Field: A Quantitative Analysis. In *Grey Exploitations in the 21st Century: Second International Conference on Grey Literature*, pp. 157-168. Amsterdam, The Netherlands; TransAtlantic GreyNet.
- Digital Library Federation. [2004] Digital Library Federation: Digital Library Documents. Available at: <http://www.hti.umich.edu/cgi/b/bib/bib-idx?c=dlf>
- Drake, Miriam. 2004. Institutional Repositories: Hidden Treasures. *Searcher* 12(5): 41-45.
- Edwards, Phillip M. 2004. Collection Development and Maintenance across Libraries, Archives, and Museums: A Novel Collaborative Approach. *Library Resources & Technical Services*, 48(1): 26-33.
- Farace, Dominic J. 1998. Foreword. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information*, pp. iii. Amsterdam, The Netherlands; TransAtlantic GreyNet.
- Gelfand, Julia M. 2000. Grey Literature Poses New Challenges for Research Libraries. *Collection Management* 24(1/2): 137-147.
- Gokhale, Pratibha A. 1998. Grey Literature Varieties – Definitional Problems. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information*, pp. 259-273. Amsterdam, The Netherlands; TransAtlantic GreyNet.
- Herring, Susan Davis. 2002. Use of Electronic Resources in Scholarly Electronic Journals: A Citation Analysis. *College & Research Libraries* 63(4): 334-40
- Intner, Sheila. 1996. The Ostrich Syndrome: Why Written Collection Development Policies are Important. *Technicalities* 16(7):1 ,9-10.
- Kyrillidou, Martha. 2000. Research Library Trends: ARL statistics. *Journal of Academic Librarianship* 26(6): 427-436.
- Latham, Joyce. 2002. To Link, or Not to Link. *Library Journal Net Connections* Spring:20-22.
- Lee, Hur-Li. 2003. Collection Development as Social Process. *Journal of Academic Librarianship* 29(1):23-31.
- Lehman, Heather & Webster, Janet G. 2004. Tsunami Digital Library Needs Assessment: Final Report to Oregon State University Libraries and the Northwest Alliance for Computational Science & Engineering. *OSU EECS Technical Report* 04-40-04. 24 pp. Available at: <http://osulibrary.oregonstate.edu/staff/websterj/webcv.htm>
- Mack, Daniel C. 2003. Introduction: Changing Collections, Changing Policies. In Mack (ed.), *Collection Development Policies: New Directions for Changing Collections*. Pp.1-2. Binghampton, NY: Haworth Information Press.
- Peters, Thomas A. 2001. That's the Big Deal? *Journal of Academic Librarianship*. 27(4):302-304.
- Russell, Beth. 2004. Special Collections Cataloging at a Crossroads: A Survey of ARL Libraries. *Journal of Academic Libraries* 30(4): 295-303.
- Snow, Richard. 1996. Wasted Words: The Written Collection Development Policy and the Academic Library. *Journal of Academic Librarianship* 22(3):191-4.
- Spohrer, James H. 2003. The End of an American (Library) Dream: The Rise and Decline of the Collection Development Policy Statement at Berkeley. In Mack (ed.), *Collection Development Policies: New Directions for Changing Collections*. Pp.33-47. Binghampton, NY: Haworth Information Press.
- Wilson, Myoung C. and Hendrik E.. 1996. Collection Development in the interdisciplinary context. *Journal of Academic Librarianship* 22: 195-200
- White, Gary W. & Crawford, Gregory, A. 1997. Developing an Electronic Information Resources Collection Development Policy. *Asian Libraries* 6(1/2): 51-56.

How to get a French doctoral thesis, especially when you aren't French

P. Paillassard, J. Schöpfel, and C. Stock
INIST / CNRS

Abstract

In 1985 the French government created a unique circuit for the dissemination of doctoral theses: References went to a national database "Téléthèses" whereas the documents were distributed to the university libraries in microform. In the era of the electronic document this French network of deposit of and access to doctoral theses is changing. How do you discover and locate a French thesis today, how do you get hold of a paper copy and how do you access the full electronic text? What are the catalogues and databases referencing theses since the disappearance of "Téléthèses"? Where are the archives, and are they open? What is the legal environment that rules the emerging structures and tools?

This paper presents national plans on referencing and archiving doctoral theses coordinated by the government as well as some initiatives for creating full text archives. These initiatives come from universities as well as from research institutions and learned societies. "Téléthèses" records have been integrated in a union catalogue of French university libraries SUDOC. University of Lyon-2 and INSA Lyon developed procedures and tools covering the entire production chain from writing to the final access in an archive: "Cyberthèses" and "Cither". The CNRS Centre for Direct Scientific Communication at Lyon (CCSD) maintains an archive ("TEL") with about 2000 theses in all disciplines. Another repository for theses in engineering, economics and management called "Pastel" is proposed by the Paris Institute of Technology (ParisTech), a consortium of 10 engineering and commercial schools of the Paris region.

(* See glossary at the end of the paper)

What is a French doctoral thesis?

Considered as scientific publications, French doctoral theses constitute an important part of scholarly communication. Following scientometrics, they represent 10-20% of indexed academic research in STM (OST* 2002).

Theses are often the result of 3-4 years of research. At the same time they are an administrative document necessary to obtain the doctoral degree. In some disciplines they are regarded as a result of teamwork and appear in the list of publications of the research laboratory (Mermet et al. 1998).

French universities are autonomous; each one delivers its own degrees and preserves the theses in its library. In the past, before 1985, the graduate student had to deposit a certain number of copies that varied according to local rules (30-180). There are more than 100 universities in France, each one with one or more catalogues and with a specific logic of preservation and supply. Furthermore, academic communities – sciences, humanities, medicine, law etc. – hold different views and have different practices and traditions. And last not least, local autonomy and responsibility are "counterbalanced" by a national framework structure, the French interlibrary loan network.

So, how find a French thesis? And once found, how get it? The following communication tries to give some practical hints and perspectives, imbedded in a larger description of the development of the production, processing and preservation of French doctoral theses and an overview of the principal actors, catalogues and databases.

First steps to improve access to this type of grey literature (1985-2000)

The French government published in 1985 a decree that regulated and improved the deposit and dissemination of doctoral theses. These rules have been applied until 2001. The main principles:

- Guarantee the deposit of the doctoral thesis.
- Harmonize the number of copies to submit.
- Facilitate the identification and availability of the documents.
- Move the format of preservation and dissemination from paper to microfiche (gain of shelf-space, easier access).

The result of the 1985 decree was the creation of a "four-level national network". Each university had to create a special service for doctoral theses ("service de doctorat"). Two institutions (ANRT*) in Lille and Grenoble transformed the print originals into microfiches. Three input centres (INIST* for sciences and

technology) centralized the creation of bibliographic records from the registration form. Finally, all records were loaded into a national online database called Téléthèses*.

Deposit and dissemination 1985-2000

Three weeks before the date of defense the candidate fills in two copies of a registration form and submits several print copies of his thesis at the "service de doctorat": one copy for each member of the jury, and three copies for the library.

The registration form contains personal, administrative and bibliographic data (including abstract and keywords with, in later years, English title and abstract) and is used for the examination process as well as for recording in the national database.

The jury may ask for modifications of the thesis to be finished within three months after the date of defense. Once the final official version submitted, the university president authorizes its reproduction and dissemination.

The print copies and registration forms are transmitted to the university library. The registration form is sent to one of the three input centres.

If authorized for reproduction, a print copy is sent to one of the national theses reproduction services (ANRT) that produces a microform version. All university libraries and some other academic institutions receive a copy on microfiche. The students' guide mentions an average dissemination of 200 microform copies per dissertation (Ministère 1994).

If the dissertation is published, the graduate student must deposit 10 sample issues at the university library (30 if the student received funding for the publication). In this case, the thesis is not converted into a microform.

French dissertations are not deposited at the National library (BNF*), and they are not included in its national bibliography. An ISBN is only attributed if the dissertation is published.

Referencing 1985-2000 – from print bibliography to online catalogue

The French ministry for education not only organized the submission of dissertations but improved its referencing as well. The registration forms were sent to three national input centres following the scientific subject (social sciences and humanities, including economics and law; medicine; and sciences).

In addition to an annual print bibliography "Inventaire des thèses" divided into three sections, a national online database was created in 1986. This new database "Téléthèses" was hosted on a university server and accessible through "Minitel", a very popular Videotex online service launched in France in 1982 but inaccessible from foreign countries.

Records in the online database referred to dissertations going back to 1972 for sciences, social sciences and humanities, to 1983 for medicine and pharmacology and 1990 for veterinary sciences.

Each record contained minimal bibliographic data, abstract and keywords in French and for a part in English. Authority lists were used for the university, type of degree and scientific domain. From 1986 on the "service de doctorat" attributed a national identification number that was included in the database record.

Between 1995 and 2003 the Téléthèses database was also published in a CD-Rom version called Docthèses*, making the database available to foreign countries. The following table contains the number of French doctoral theses referenced by Docthèses between 1993 and 2002:

Year	Number	Humanities and social sciences	Sciences	Medicine and veterinary sciences
2002 (*)	10720	3209	?	?
2001 (*)	8922	2612	6310	?
2000 (*)	10663	3572	4977	2113
1999	17826	3530	6011	8285
1998	18444	3398	6655	8391
1997	19886	3637	7083	9166
1996	19542	3460	7181	8901
1995	18449	3019	6422	9008
1994	19344	3307	6613	9424
1993	18813	2844	6335	9634

Table 1: Theses referenced in the "Docthèses" CD-Rom database (1993-2002)
(*) 2000-2002 are transition years and the number of theses is not complete

In 2000 the Téléthèses database moved from Videotex to a web server hosted by ABES*. At the same time, all records were loaded into the new national academic union catalogue SUDOC*.

From 2001, the university libraries started to create "their" records directly in the SUDOC, and the online and CD-Rom databases disappeared. The SUDOC catalogue contains today more than 500 000 theses.

The INIST online database Article@INIST* contains nearly 100 000 theses, most of them from science and technology.

Critics of the 1985-2000 system

The 1985 decision facilitated recording and availability of French theses. Nevertheless, some critics arose especially from library professionals:

- Workload: University libraries couldn't download the records from the database, but had to key them again for their own catalogue.
- Incomplete information: Especially in humanities and social sciences, librarians wanted to increase reference quality by adding national subject headings (RAMEAU*).
- Delays: The interval between the date of defence and the moment the records were integrated into the union catalogue was sometimes rather long.
- Supply price: The price of dissemination of theses through print copies from microforms was generally considered as too high.

In 1996 the centralized input of records for theses in social sciences and humanities was replaced by direct input from each library.

Finally these critics, together with the development of electronic theses and dissertations (ETDs) in France and other countries (see for instance Friend 1998 for the UK, Jin 2004 for China, Suleman & Fox 2003 for the international Networked Digital Library of Electronic Theses and Dissertations, Rutledge 1994), lead from 1998 on to a radical change in national politics in favour of a national solution for electronic theses (see Okret-Manville 2002).

From print to electronic format (1998-2004)

The French ministry of education developed since 1998 the project of a national server for electronic theses and dissertations (Okret 1998). This project was meant to substitute the 1985 network. Its fundamental assumptions:

- The centralized structure is preserved.
- In the heart of the system, the national academic union catalogue, SUDOC.
- Each ETD record in the SUDOC is linked to the full-text (URL link from the 856 field).
- Each ETD is archived on a local server (university).
- A national backup server contains part or all of French ETDs in PDF.

The 1998 project already referred to three problems: format for preservation and transmission, protection of intellectual property, need of investment for the national server and local archives.

In the following year, the ministry created a commission to prepare the technical aspects of this project. The conclusions – a unique model with the same software and procedures for all universities – were published in a report (Jolly 2000). In 2001, the ministries of education and research published guidelines for students and universities with detailed recommendations for digital editing, archiving and supply that were based on the program “Cyberthèses” developed by the University Press of Montreal and the university of Lyon-2:

- Edition: native format compatible with RTF or TeX, with a common style sheet in order to structure the whole document.
- Preservation: conversion into XML.
- Format of dissemination: PDF, HTML or XML.
- Each university records its theses in its own catalogue and in the national union catalogue (Unimarc format).
- Metadata: the 2000 report included a data model derived from the Dublin Core. In the future, metadata harvesting by the ABES from local academic servers was supposed to substitute the traditional recording. The metadata then would be reformatted into Unimarc records for the SUDOC catalogue.
- The full text is archived on a local university server; a backup copy is preserved on a national server run by the CINES*. The university library has to preserve a print copy.
- Supply: by the university, online access and/or through interlibrary loan (print copy).

A number of articles and communications were published to inform about the project and encourage local progress (see Boyer et al. 2001, Okret-Manville 2002).

The 2004 reality: diversity, problems, and perspectives

Three years later, the French ETD landscape is all but homogeneous. The results of the government initiative seem disappointing. The development and implementation of national software and services progress slower than planned. A study ordered by the ministry (spring-summer 2004) totals only 360 ETDs in conformity with the governmental guidelines, less than 5% of the annual number of theses (Six&Dix 2004).

In the same time, a growing number of alternative, more or less successful local initiatives, academic networks and open archives give access to more than 4000 ETDs.

The reasons for this paradoxical situation are various.

Up to now, neither the government nor any other institution had enough coercive or persuasive force to impose a unique model for ETDs. Perhaps this “unique model” is simply unrealistic and non-adapted to the heterogeneous needs, behaviours and traditions of scientific and academic communities.

Another reason is financial: government funding was centred rather on the development of the national union catalogue and access to online resources (“big deals” by the academic consortium) than on ETDs. Without specific financial support by the government, the local investment by universities was often limited. The Ministry’s initial evaluation of human and budget resources (one librarian and 10 000 euros for the processing of 100 ETD/year) was too optimistic.

Underestimated was also the need for new technical knowledge and procedures, training of graduate students and investment for new soft- and hardware (see Laloë 2003). Generally, the technical requirements were considered as too complicated, both from academics and librarians and from students.

Together with partly lacking motivation of universities and students, these problems significantly slowed down the development of the national networked digital library.

But just as in other countries, some universities and academic communities started to develop their own and often less complicated ETD solutions, comparable for instance to Virginia Tech or ETH Zürich (see Jutzi & Keller 2001). Probably the initial project based on a unique model will shift to a modular network based on mixed deposit (print/native format), PDF/XML archiving and PDF/HTML supply (see Six&Dix 2004) and take into account alternative solutions comparable for instance to the ProQuest/UMI online submission system (see Cox & Barbosa-Jerez 2004).

French ETD archives in 2004

The following chapter offers an overview of the seven most representative digital archives¹ that give free access to French ETDs. These archives were developed since 1997 and 1998 by French universities, engineering schools, national institutes and the CNRS*. Figures and data are from November 5th, 2004. The appendix contains more detailed information for each of these ETD archives.

(a) Physics, mathematics, chemistry and engineering sciences

- **Cithier:** produced by INSA* Lyon with 174 theses in the engineering field.
- **Pastel:** produced by the Paris Institute of Technology (ParisTech with 10 independent engineering schools). Pastel contains 341 dissertations with online access to the full text.
- **MathDoc*:** developed by the university of Grenoble-1 and the CNRS. MathDoc is one of the oldest French archives with more than 1000 theses in mathematics.
- **INRIA*:** the INRIA archive gives access to more than 1000 theses in computer science and control.

(b) Multidisciplinary archives

- **Cyberthèses:** a common project between Canadian and French universities (Montreal, Lyon), gives for example access to 366 multidisciplinary ETDs for Lyon 2 University.
- **Grisemine:** is produced by the university of Lille-1. Its 396 theses cover sciences, technology and social sciences.
- **TEL:** created by the CCSD* and MathDoc*. It is today the most comprehensive French archive with 2292 ETDs in full text, covering all domains but mostly physics, mathematics and engineering sciences.

(c) Typology of archives

Four different types of archives can be distinguished, even if these types are not exclusive:

- The **institutional archive** contains all theses of one (INRIA) or more than one structure (Pastel).
- The **domain-specific archive** gives access to ETDs from different establishments but of the same scientific domain (MathDoc).
- The **collaborative or multi-side archive** offers facilities to different structures (International program Cyberthèses).
- The **multi-type archive** contains ETDs but also other academic literature - preprints, conference papers, courseware and so on (Grisemine).

These criteria are not exclusive; a given archive can belong to more than one category as for instance Grisemine that is also the institutional archive of the university of Lille-1.

The most frequent type seems to be the collaborative or multi-side archive. The cooperation can be realized on different levels:

- **Management and administration:** Cyberthèses is co-managed mainly by the universities of Montreal and Lyon-2 and a French foundation for information highways (Fonds Francophone des Inforoutes).
- **Coverage:** Multilingual research interfaces are more and more frequent. TEL offers French, English and German versions. ETDs are in different languages and come from different European, African and American countries.

This willingness to cooperate is reinforced by the use of metadata harvesting through the OAI PMH protocol and the use of open source software. Pastel, TEL and Cyberthèses are declared OAI. Grisemine as an institutional archive for French grey literature is moving from a proprietary system to the Dspace software from the MIT.

(d) Other services and functionalities

Some archives offer more than full text access and include special and complementary services, for instance:

¹ Digital archives does not necessarily mean open archive. By digital archive we understand platforms, institutional servers, repositories who give access to ETD full text in a permanent way.

- Complete editorial chain: Cyberthèses proposes a complete editorial chain called "Cyberdocs" from a document model to the conversion into a fully structured XML document using TEI lite DTD. Discussion lists and downloadable tools complete the offer.
- Links to online services: MathDoc offers links to different special portals and online services such as the Zentralblatt-MATH (FIZ Karlsruhe), the MathSciNet (American mathematical society), or Springer Link.
- Online submission: TEL and Pastel permit online submission by the author (self-archiving). Even so, in most cases the institution before making them available controls metadata and documents.

Up to now, we found no study on usage patterns of the different French archives and systems comparable to Zhang, Lee & You 2001 for the Korean KISTI system.

Detailed aspects can be found in the individual presentations of each archive (see appendix).

(e) Perspectives

More and more universities and organizations encourage the submission of "their" ETDs to the CCSD archive TEL. Others "paste" them directly to this archive, sometimes by automatic procedures, for instance the IMAG (Institute of Applied Mathematics and Computer Science at Grenoble), IN2P3 (Institute of Nuclear Physics and Particle Physics), and ParisTech with its Pastel archive.

Presently, TEL has become the most important ETD archive in France, with the highest number of participating organizations and ETDs. CCSD tries to increase its coverage for humanities and social sciences and life sciences, prepares the retro-digitisation of print theses and intends to place the archive under international control in order to ensure permanent access.

It is possible that in the next future the French government will confirm this reality and that the TEL archive will officially become the central national ETD archive.

ETD metadata: towards a national metadata scheme

Following the results of the Jolly report, AFNOR*, the French standardization organization charged an expert group to define the metadata required with the national deposit of ETDs.

Based on the Dublin Core, the national metadata scheme is written in XML. It offers a match to Unimarc fields of the SUDOC records. It will also be compatible with the OAI protocol for metadata harvesting.

Compared to the traditional paper registration form, the new scheme is richer. In addition to the "traditional" bibliographic metadata, the scheme includes several administrative data as well as information related to the life cycle of ETDs and to the rights management. Other data are optional, for instance the name of the research laboratory.

Legal aspects

In the 1980's a thesis was considered as a *university* document that should be disseminated as widely as possible. According to their examination regulations, the universities considered the jury's authorization sufficient for dissemination.

With the appearance of ETDs and the evolution of the author's rights, a dissertation is no longer seen as a "university document" but as a work subject to intellectual property rights.

Today the explicit authorization by the author of the thesis (= copyright holder) is necessary for the electronic dissemination, in addition to the jury's decision. This authorization should be requested when the thesis is submitted (Jolly 2000). Furthermore, some universities ask for a declaration of conformity between electronic and print version and/or between the native deposit format and the XML version (Six&Dix 2004).

Some universities (Metz, for instance) already started to search for their former graduate students in order to obtain an authorization for retro-digitisation and online access of older theses.

At present, the impact of the ETD policy on scholarly publication (see Seamans 2003) isn't addressed in France but may be a reason for the lack of motivation of some candidates to accept electronic submission and free availability of their thesis through Internet posting.

Conclusion: some practical tips to search and order a French thesis

From a clearly structured network in the 80s and 90s with defined roles, actors and services, the French dissertation landscape has changed into a heterogeneous mixture of national structures and local initiatives. This may be characteristic for a transitional period from a traditional "print circuit" to a networked digital library of ETDs. In the meantime, searching for French theses has to adopt a double strategy, based on an interrogation of the academic union catalogue SUDOC and a web-based search in ETD archives and repositories.

How to find a thesis in the SUDOC:

- Choose the "Extended Search" interface.
- De-select all types of publication except for dissertations.
- Choose or select a subject.
- Limit the publication year or range.
- Add keywords with the index "subject words".
- For formal information select the index "dissertation note". This index contains formal information about the type of theses, the domain, the university and the date.

Each bibliographic record in SUDOC is linked to a holding record that lists the university libraries in possession of the document, with details on loan/copy conditions (PEB*).

In some special cases it is difficult or impossible to obtain a thesis referenced in the SUDOC:

(1) Confidential theses are referenced in the databases or university catalogues, but are not available. The principal reasons for confidentiality are:

- The research has been conducted on a subject where patents have been submitted.
- The author plans to publish his work commercially.

If the confidentiality is time-limited, the document becomes available after this period.

(2) The jury/commission may ask the candidate to revise parts of his dissertation. If this isn't done then the thesis may not be disseminated officially and be excluded from microform reproduction. Even if it could eventually be retrieved from a personal website, its scientific value should be considered with prudence.

(3) "Thèses d'exercice" in medicine normally are not reproduced in microform. They are available at the student's university and at the central library for medicine in Paris (BIUM*) where they can be retrieved through the BIUM catalogue.

Print copies from French theses can also be ordered directly via the INIST document supply service.

The Lille ANRT offers a service called "Thèse à la carte" where theses can be searched by subject or domain and ordered in book format; presently, the ANRT catalogue contains roughly 4500 theses.

Even if the SUDOC catalogue remains the point of access to all French theses in print format, ETDs must be searched in the different local and networked archives and databases to obtain full text access, since the SUDOC still offers a rather small number of records with hyperlinks to documents.

The search for a French ETD can start in some digital libraries or portals that offer updated selections of web links to repositories and archives.

Web links to ETD archive information:

Agence bibliographique de l'enseignement supérieur, Thèses [Online]. -
<<http://www.abes.fr/abes/DesktopDefault.aspx?tabid=223>>

Bibliothèque Nationale de France, Thèses francophones en ligne [Online]. -
<http://signets.bnf.fr/html/categories/c_011theses_fra_ligne.html>

Ecole nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB), Sibel. Thèses [Online]. -
<<http://sibel.enssib.fr/index.php?m=c&c=479>>

Maison des Sciences de l'Homme -Alpes, Thèses [Online]. -
<<http://www.msh-alpes.prd.fr/veille/actualiterecherche.htm>>

Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche, Thèses en ligne [Online]. -
<<http://www.educnet.education.fr/dossier/rechercher/these1.htm>>

Another way is to search directly in the ETD archives (see appendix) or on the universities' websites and catalogues. Nevertheless, in spite of these initiatives and services, searching French ETDs still remains a more or less difficult task.

But this difficulty isn't it just the main feature of grey literature? And as mentioned above, it reflects the transition from a well-organized "print network" to a new and open structure where centralized services such as the SUDOC or a national ETD archive probably hosted by the CINES or the CCSD (TEL), together with other French or European federate sites and portals (Cyberthèses, INIST, the future SIGLE gateway to grey literature) will facilitate search and access to French ETDs. The ideas and projects exist; their realization is a matter of public funding, resources and time.

References

- Boyer, G., Coulouma, E. and Lagarrigue, C. (2001), Thèses. -In: Arabesques. -No. 22 (2001), pp. 4-5.
- Cox, F.M. and Barbosa-Jerez, M. (2004), Gleanings from the 7th International Symposium on Electronic Theses and Dissertations. -In: Library Hi Tech News. -No. 8 (2004), pp. 10-12.
- Claerebout, M.-F. (2003), Grisemine, a digital library of grey university literature. -In: GL5 Conference Proceedings, Amsterdam, 4-5 December 2003. -pp.27-31
- Friend, F.J. (1998), Brief communication: UK theses online?. -In: Interlending & Document Supply. -Vol. 26, no. 4 (1998), pp. 175-177.
- Jin, Y. (2004), The development of the China Networked Digital Library of Electronic Theses and Dissertations. -In: Online Information Review. -Vol. 28, no. 5 (2004), pp. 367-370.
- Jolly, C. (2000), Rapport sur la diffusion électronique des thèses [Online]. Ministère de l'éducation nationale. Direction de l'enseignement supérieur. Sous-direction des bibliothèques et de la documentation. -
<<http://www.sup.adc.education.fr/bib/>>
- Jutzi, V., Keller, A. (2001), Dissertationen Online an der ETH-Bibliothek Zürich. -In: Bibliotheksdienst. -Vol. 35, no. 3 (2001), pp. 306-312.
- Laloë, F. (2003), Une ambition universitaire légitime: construire les bibliothèques numériques [Online]. -
<http://web-linux.univ-lyon2.fr/ed_uni/article.php3?id_article=2>
- Mermet, J.-M., Prudhomme, B., Joly, M., Pinon, J.-M. (1998), CITHER: un modèle de bibliothèque numérique de mémoires de thèses. -In: Document numérique. -Vol. 2, no. 3-4 (1998), pp.51-72.
- Ministère de l'Enseignement Supérieur et de la Recherche (1985), Arrêté du 25 septembre 1985 relatif aux modalités de dépôt, signalement et reproduction des thèses soutenues ou travaux présentés en soutenance en vue du doctorat [Online] -
<<http://www.sup.adc.education.fr/bib/Acti/These/textregl.htm#25sept>>
- Ministère de l'Enseignement Supérieur et de la Recherche (1994), Le signalement et la valorisation de la thèse. Guide du candidat au doctorat.
- Ministère de l'Education Nationale & Ministère de la Recherche (2000), Circulaire du 21 septembre 2000: Diffusion électronique des thèses. Circulaire no. 2000-149 [Online]. -In: Bulletin Officiel de l'Education Nationale. -No. 34, 28 September 2000. -
<<http://www.education.gouv.fr/bo/2000/34/sup.htm>>
- Ministère de l'Education Nationale & Ministère de la Recherche (2001), Guide pour la diffusion et l'archivage électroniques des thèses à l'attention des établissements d'enseignement supérieur [Online] -
<<http://www.sup.adc.education.fr/bib/>>
- Okret, C.(1998), Un élément de la rénovation du dispositif de diffusion des thèses françaises : un serveur national de thèses numérisées. In: Lettre d'information du CNUC. -No. 67 (1998).
- Okret-Manville, C. (2002), Diffusion des thèses. -In: Arabesques. -No. 26 (2002), pp.7-8.
- OST - Observatoire des Sciences et Techniques (2002), Indicateurs bibliométriques des institutions publiques de recherche. Année 1997 - hors sciences humaines et sociales. Essai méthodologique [Online]. -
<<http://www.obs-ost.fr>>
- Rutledge, J. B. (1994), European Dissertations: production, access, use. -In: Collection management. -Vol. 19, no. 1-2 (1994), pp. 43-67.
- Seamans, N.H. (2003), Electronic theses and dissertations as prior publications: what the editors say. -In: Library Hi Tech. -Vol. 21, no. 1 (2003), pp. 56-61.
- Six&Dix (2004), Etude de la mise en œuvre du dispositif national de diffusion des thèses par voie électronique dans les Etablissements d'enseignement supérieur et de recherche. Note de synthèse [Online]. -
<<http://www.sup.adc.education.fr/bib/>>
- Suleman, H. & Fox, E.A. (2003), Levering OAI harvesting to disseminate theses. In: Library Hi Tech. -Vol. 21, no. 2 (2003), pp. 219-227.
- Zhang, Y., Lee, K. and You, B.-J. (2001), Usage patterns of an electronic theses and dissertations system. -In: Online Information Review. -Vol. 25, no. 6 (2001), pp. 370-377.

Glossary

ABES: Agence Bibliographique de l'Enseignement Supérieur (operating agent of the French academic union catalogue and ILL system): <http://www.abes.fr>

AFNOR: Association Française de Normalisation (French standardisation organisation): <http://www.afnor.fr/>

ANRT: Atelier National de Reproduction des Thèses (national service for the reproduction of theses). The Lille ANRT is hosted at the Charles de Gaulle university at Lille-3: <http://www.anrtheses.com.fr>

Article@INIST: Online database with over 8 million references of articles and monographs for document supply: <http://services.inist.fr/public/eng/conslt.htm>

BIUM: Bibliothèque Interuniversitaire de Médecine (largest medical library in France and most important supplier in the academic sector): <http://www.bium.univ-paris5.fr/>

BNF: Bibliothèque Nationale de France (French national library): <http://www.bnf.fr/>

CCSD: Centre pour la Communication Scientifique Directe (CNRS Centre for Direct Scientific Communication): <http://ccsd.cnrs.fr/>

CINES: Centre Informatique National de l'Enseignement Supérieur (national academic computer and data-processing centre, called CNUSC until 1999): <http://www.cines.fr/>

CNRS: Centre National de la Recherche Scientifique (French National Research Organisation): <http://www.cnrs.fr/>

INIST: Institut de l'Information Scientifique et Technique (CNRS institute for scientific and technical information): <http://www.inist.fr>

INRIA: French national institute for research in computer science and control: <http://www.inria.fr>

INSA: Institut National des Sciences Appliquées de Lyon (one of the top French engineering universities): <http://www.insa-lyon.fr/>

MathDoc: French network for documentation in mathematics and server for the management of ETDs run by the university of Grenoble-1 and the CNRS: <http://math-doc.ujf-grenoble.fr/Theses>

OST: Observatoire des Sciences et des Techniques (Observatory for Science and Technology, production of quantitative indicators for S&T activity): <http://www.obs-ost.fr/fr/>

PEB: Prêt entre bibliothèques (French academic interlibrary loan)

RAMEAU: Répertoire d'autorité-matière encyclopédique et alphabétique unifié (academic subject headings authority list): <http://rameau.bnf.fr>

SUDOC: Système Universitaire de Documentation (academic union catalogue of serials and monographs): <http://www.sudoc.abes.fr/>

All websites visited in November 2004.

Appendix

The following sheets contain for each ETD archive four sets of data:

- General information: content, domains, coverage (anteriority) and number of ETDs.
- Technical information: software and formats.
- Archiving: self-archiving and legal aspects.
- Institutional information: institution, website, contact, complementary information.

All data were collected between July and November 2004.

CITHER**General information**

URL	http://csidoc.insa-lyon.fr/these/
Content	Theses from INSA
Type	Institutional
Language	French
Domains	Engineering sciences
Input	INSA Lyon
First release	1997
Number of theses	174
First thesis	1996

Technical information

Platform/software	CEN (DELPHI)
Deposit format	Word, Latex, PostScript
Viewing format	PDF
Storage format	Word, PDF

Archiving

Self-archiving	No
Dissemination	Authorization by author and university
Removal of thesis	Allowed

Institutional information

Institution	Centre de documentation scientifique Doc'INSA
Website	http://docinsa.insa-lyon.fr/
Contact	cither@insa.insa-lyon.fr

CYBERTHESES**General information**

URL	http://cyberdocs.univ-lyon2.fr/
Content	ETDs from Cyberthèses members
Type	Collaborative
Language	Website in French and in English
Domains	Multidisciplinary
Input	Universities of Lyon-2, Montreal, Chili and several institutions from other countries
First release	1998
Number of theses	366 (Lyon 2)
First thesis	1996

Technical information

Platform/software	Cyberdocs (platformGPL), open source software
Editorial chain	Conversion of structured word documents in XML with the TEI Lite DTD
Metadata format	Dublin Core and specific elements
OAI-PMH	Yes
Deposit format	Latex, Word, WP, Lotus...
Viewing format	HTML, XHTML, PDF
Storage format	XML

Archiving

Self-archiving	No
Dissemination	University agreement requested

Institutional information

Institutions	University of Montreal University of Lyon-2
Websites	http://www.umontreal.ca/ http://www.univ-lyon2.fr/
More information	CyberThèses, Agence Intergouvernementale de la Francophonie, FFI, Cyberdocs: structure and publish electronic documents [Online]. - http://sourcesup.cru.fr/cybertheses/ University of Lyon 2, Cyberthèses [Online]. - http://mirror-fr.cybertheses.org/en/cybertheses.php
Contact	Jean-Paul DUCASSE : Jean-Paul.Ducasse@univ-lyon2.fr Guylaine BEAUDRY : guylaine.beaudry@umontreal.ca

GRISEMINE**General information**

Content	French grey literature
URL	http://cridoc.univ-lille1.fr/default.asp?bustl/grisemine
Type	Institutional, collaborative, multi-type
Language	French
Domains	Sciences, technology and social sciences
Input	University of Lille-1
First release	2001
Number of theses	396 (full text or abstracts)
First thesis	End of 19th century
Others documents	Preprints, conference papers, etc
Total documents	1039

Technical information

Platform/software	Is moving from a proprietary software to Dspace
Metadata format	Dublin Core and specific elements
Deposit format	ASCII, Word, RTF, PDF, PostScript, HTML
Viewing format	PDF
Storage format	PDF

Archiving

Self-archiving	No
Dissemination	Authorization by author and university
Removal of thesis	Allowed

Institutional information

Institution	University of Lille-1
Website	http://ustl1.univ-lille1.fr/projetUstl/
More information	Claerebout, M.-F., Marino, J.-B., Grisemine, une bibliothèque numérique de littérature grise universitaire [Online]. Amsterdam : GL5, 4-5 December 2003. - < http://archivesic.ccsd.cnrs.fr/sic_00000892.en.html >
Contact	Marie-France CLAEREBOU : marie-france.claerebout@univ-lille1.fr

INRIA**General information**

Content	Research reports, technical reports, thesis edited by INRIA
URL	http://www.inria.fr/rrrt/index.fr.html
Type	Institutional, multi-type
Language	English, French
Domains	Communication and information science, computer science, automation
Input	INRIA
Number of theses	About 1000
First thesis	1985

Technical information

Platform/software	Alta Vista Search Intranet (will change soon)
OAI-PMH	In project- in cooperation with CCSD
Deposit format	LaTeX...
Viewing format	PDF, PostScript

Institutional information

Institution	INRIA
Website	http://www.inria.fr/index.fr.html
More information	INRIA, Annual report 2003 [Online]. - < http://www.inria.fr/inria/rapportannuel/ran.fr.html >
Contact	webmaster@inria.fr

MATHDOC**General information**

Content	Online access to resources in mathematics
URL	http://math-doc.ujf-grenoble.fr/Theses
Type	Collaborative, domain-specific
Language	Website in English and French, documents in different languages
Domain	Mathematics
Input	Universities (institutes of mathematics), CCSD
First release	1998
Number of theses	1000 (full text and abstracts)
First thesis	1990
Others documents	Preprints, journals, books, digital library of old documents

Technical information

Platform/software	Harvest, OpenResolver (Open URL), open source software
Formats	See TEL

Archiving

Self-archiving	Self-archiving via TEL
Dissemination	See TEL
Removal of thesis	Allowed

Institutional information

Institution	Cellule MathDoc CNRS University of Grenoble-1
Website	http://www-mathdoc.ujf-grenoble.fr/ http://www.ujf-grenoble.fr/ujf/fr/welcome.phtml
More information	Cellule MathDoc, Comité de pilotage: 29 septembre 2003 [Online]. - < http://www-mathdoc.ujf-grenoble.fr/Activites/Rapports/CP-CMD_2003_web.pdf > INIST-CNRS, La Cellule MathDoc [Online]. - < http://www.inist.fr/openaccess/article.php3?id_article=30&var_recherche=mathdoc >
Contact	accueil@mathdoc.ujf-grenoble.fr

PASTEL**General information**

Content	Theses from ParisTech
URL	http://pastel.paristech.org/
Type	Institutional, collaborative
Language	French
Domains	Engineering sciences
Input	Paris Institute of Technology includes 10 engineering schools
First release	2002
Number of theses	341
First thesis	1994

Technical information

Platform/software	Eprint 2.2.1
OAI-PMH	Yes
Metadata format	Dublin Core and specific elements
Deposit format	PDF, PostScript, HTML, MS Word...
Viewing format	HTML, PDF, PostScript
Storage format	PDF, HTML, Text (ASCII)...

Archiving

Self-archiving	Yes
Dissemination	Authorization by author and university
Removal of thesis	Allowed

Institutional information

Institution	ParisTech
Website	http://www.paristech.org/anglais/index.html
Contact	admin@rilk.com

TEL**General information**

Content	Multidisciplinary theses server
URL	http://tel.ccsd.cnrs.fr
Type	Collaborative
Language	Website in English, German and French, theses in different languages
Domains	Multidisciplinary but mostly physics, mathematics and engineering sciences
Institutions	French universities, universities from other countries
First release	2001
Number of theses	2292
First thesis	1990

Technical information

Platform/software	Eprint 1.1
Editorial chain	No
OAI-PMH	Yes
Deposit format	(La)TEX, RTF, Word, PostScript, EPS, JPEG, HTML, PDF
Viewing format	(La)TEX, Word, HTML, PostScript, PDF, XML
Storage format	(La)TEX, Word, XML

Archiving

Self-archiving	Yes
Dissemination	Authorization by author, often also by university
Removal of thesis	Allowed

Institutional information

Institution	CCSD Cellule MathDoc
Website	http://www.ccsd.cnrs.fr http://www-mathdoc.ujf-grenoble.fr/
More information	Laloë, F.: franck.laloe@lkb.ens.fr
Contact	adminTEL@ccsd.cnrs.fr

Author Information

Mrs. Pierrette Paillassard,
 Librarian for Theses and Dissertations
 Institut de l'Information Scientifique et Technique, INIST/CNRS
 2, allée du Parc de Brabois
 54514 Vandoeuvre-Les-Nancy Cedex
 France

EML: Pierrette.Paillassard@inist.fr
 URL: <http://www.inist.fr/>

ORNL – SAFEGUARDS AND SECURITY : The Global producers of the 8 radionuclides of particular security concern

James Carman, Althea Creel, Steven Milewski, James Staub and Angela Woofter
The School of Information Sciences; University of Tennessee, Knoxville - UTK

Project Report

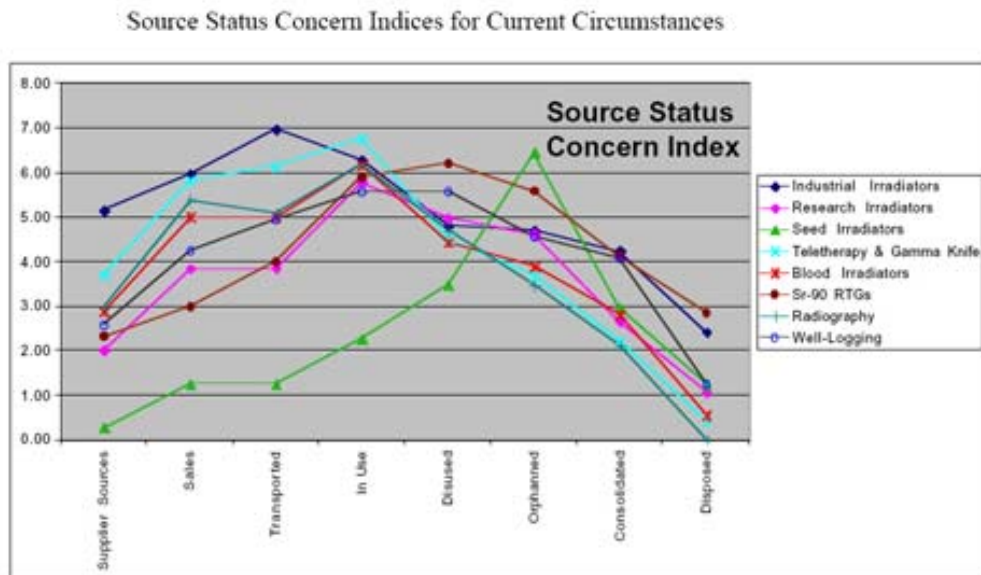
IS 566: Business Intelligence
Instructor: Dr. Kendra Albright
April 15, 2004

Introduction

The scope of this assignment was to identify research and production nuclear reactors that are producers of significant quantities of eight specific radionuclides that are of security concern: Americium-241; Californium-252; Cesium-137; Cobalt-60; Iridium-192; Plutonium-238; Radium-226; and Strontium-90, by Mr. David Lambert, the Program Co-Manager of the Material Protection, Control, and Accounting Division of Oak Ridge National Laboratory's Nuclear Nonproliferation Programs. The ultimate goal of this report is to aid in prioritizing the security risk for world-wide producers of these eight radionuclides.

Nuclear non-proliferation is important because there is a danger of societal rebels or extremist political or religious groups using certain radiological sources to create widespread panic and negative economic consequences (Tuyle et al. 2003). Lax security measures, such as legitimate users' abandonment of stockpiles resulting in "orphan sources," provide tantalizing targets for terrorists (Buxbaum 2004).

The following graph illustrates some source indexes that investigators at Los Alamos have identified as being sources of concern (Tuyle et al. 2003):



List of Countries mentioned in the Study

- Argentina
- Australia
- Belgium
- Brazil
- Bulgaria
- Canada
- China
- Czech Republic
- Egypt
- Germany
- India
- Indonesia
- Japan
- Korea (North)
- Korea (South)
- Morocco
- Netherlands
- Norway
- Romania
- Russia
- South Africa
- Sweden
- Switzerland
- United Kingdom
- United States

The eight radionuclides

The following briefs provide general information regarding each of the eight radionuclides of greatest concern (U.S. Dept. of Energy et al.):

Americium-241 (Am-241) is produced from plutonium, emitting alpha particles and gamma rays and having a half-life of 432.7 years. Among its uses are as a component in medical diagnostic devices; fluid-density, thickness, and aircraft fuel gauges; and in distance-sensing devices. All of us have come in contact with americium-241 as a component found in almost every home and business – in the ionization chamber inside a smoke detector. This radioactive isotope poses significant health risks if swallowed or inhaled. Staying in the body for years, exposure to its alpha and gamma radiation increases a person's risk of developing cancer. The primary danger from unknown exposure to americium-241 comes from industrial instruments that have been lost or stolen, or are now outside monitored control efforts, becoming "orphan sources."

Californium-252 (Cf-252) is manufactured by bombarding curium-242 with a type of helium ion. If one needs this element to use in neutron moisture gauges or for discovering gold or silver by activation analysis, one can obtain it for about \$10 per milligram through the Oak Ridge National Laboratory, but one must be quick – it has only a 55-day half-life. Exposure to californium-252 results in biological hazards.

Cesium-137 (Cs-137) results from uranium and plutonium absorbing neutrons and undergoing fission. With a half-life of thirty years, it is one of the more common industrial radioisotopes. Among its uses are in medical therapy to treat cancer, in moisture-density, leveling and thickness gauges, and in devices used to characterize rock strata. While everyone is exposed to cesium-137 daily through soil and water contaminated with atmospheric fallout, higher exposure may result in increased risk of cancer, serious burns, acute radiation sickness, and death. The typical accidental exposure results from handling what one believes to be scrap metal or through exposure from one of the measuring devices mentioned above.

Cobalt-60 (Co-60) is produced when metal structures are exposed to neutron radiation or through linear acceleration. With a half-life of 5.27 years, it is commonly used in radiation therapy by implant, in leveling gauges, x-raying welding seams to detect flaws, and for food irradiation. Prolonged exposure may result in cancer, skin burns, acute radiation sickness, and death. Depending on quantity for a particular activity, one may acquire cobalt-60 from \$1 to \$10 per curie.

Iridium-192 (Ir-192) is produced by neutron bombardment of iridium, having a 73.83-day half-life. While it emits beta and gamma radiation, its toxic properties are not yet fully determined. Exposure to iridium-192 was blamed for eight deaths from lung hemorrhages at a Moroccan construction site in 1984. It is used in brachytherapy and intracoronary radiation therapy, as well as for checking flaws in welds.

Plutonium-238 (Pu-238), with a half-life of 87.7 years, is created from the decay of uranium-238. Its radioactive decay results in a useable heat source convertible to electric power. This power may be used to sustain satellites or in very tiny amounts to fuel heart pacemakers. Nuclear weapon production and testing, accidents such as Chernobyl, and reentering satellites have produced widespread environmental contamination. Very little health risk is associated with external exposure to plutonium-238; however, ingestion or inhalation may result in increased risk of cancer and damage to the kidneys.

Radium-226 (Ra-226) is an isotope of naturally occurring radium, as uranium-238 decays. Radium-226 has the longest half-life of radium's isotopes, 1,600 years, and decays to radon-222. While people may be exposed to radium and its isotopes through various contaminated resources, the greatest health risk comes from its decayed version, radon. Once ingested or inhaled, radium usually collects in the bones and remains there for a long period of time. Again, continuous exposure to radium and its isotopes may cause cancer and other body disorders; but continuous exposure to its decayed form as radon is much more harmful to humans.

Strontium-90 (Sr-90) was discovered during nuclear experiments associated with development of the atomic bomb in the 1940s. Large amounts of strontium-90 were dispersed into Earth's atmosphere during those weapons testing activities as well as after the Chernobyl accident. Sustained exposure may result from working in scrap metal and metal melting and casting activities as well as from municipal landfill operations. Strontium-90 has a half-life of 29.1 years. With a chemical behavior resembling calcium, once ingested or inhaled, strontium-90 accumulates in the bones and teeth. This radioisotope is used as a tracer in medical and agricultural studies, and its heat generating ability provides power for long-lived, light-weight power supplies such as those used in navigational beacons, weather stations, and space vehicles, and as a component of electron tubes and industrial gauges. While useful in the treatment of eye diseases and bone cancer, ironically its dangers include bone cancer, cancer of the soft tissue near the bone, and leukemia.

As one can see, these eight radioisotopes pose both internal and external hazards, and many of them emit gamma rays: electromagnetic radiation emitted by radioactive decay and having energies in a range from ten thousand (10^4) to ten million (10^7) electron volts (Dictionary.com 2004). Only a small fraction of commercial radioactive sources pose inherently high security risks. These are of more concern because they are portable, dispersible, and more radioactive (Ferguson 2003). These isotopes are also of concern because they generally have long half-lives and pose greater long-term risks to health (David Lambert, interview, 16 March 2004).

A dirty bomb, or more formally a Radiological Dispersal Device, commonly known as an "RDD," is a device that combines conventional explosives with radiologically active materials. The detonation of the device would result in the dispersal of radioactive material over a wide area, depending on the amount of materials, the strength of the explosives used, and the prevailing weather conditions (David Lambert, interview, 16 March 2004).

Another possible method of terrorist deployment would be to place any of the isotopes such as cobalt, iridium, or radium in an area that gets a lot of repeat pedestrian traffic. Places such as subways, stairways, elevators, and meeting rooms would enable the terrorists (potentially) to repeatedly expose the same people to harmful radiation. Such placement could be very damaging over time and not need an explosive device to cause widespread harm.

Monitoring

Given the potential damage that these materials could cause, it is important to monitor the development and distribution of these radionuclides. This is particularly true with distribution, because we cannot be sure of the security measures that are in place when the radionuclides are delivered to customers throughout the world (David Lambert, interview, 16 March 2004). Questions of proper waste storage and disposal continue to be problematic in certain regions as well.

Methodology

In the client interview, Mr. Lambert stated that he was interested in which isotopes were produced by which facilities in which amounts for whom. He elaborated by saying that any reactor — whether commercial or research — could be a producer of these isotopes and be involved in some form of radioisotope sales. He also pointed out that any amount of any of the eight radionuclides produced could be an area of concern. Mr. Lambert pointed out that, if possible, we should also look for the direct recipients of the radionuclides produced. Finally, Mr. Lambert stated that he would like our findings in the form of a simple matrix.

From the information provided to us in the client interview, we began this investigation with the approach of "what, where, and for whom" and recording what we found in a database (see **Results** section below). The purpose was to identify concentrations of reactors that produce these eight radionuclides throughout the world. Since our starting reference was the International Atomic Energy Association's online Research Reactor Database, our own database and subsequent research reflect this IAEA source.

For example, the IAEA database lists reactors by country, so rather than each member of the group researching an individual radionuclide, each of us took a continent or region of the world and

searched for reactors by country within those continents or regions. Also, for consistency, our database uses the IAEA facility names and IAEA country designations where possible.

In addition to the IAEA, we turned to governmental authorities such as the US Department of Energy the US Nuclear Regulatory Commission, as well as atomic regulatory organizations in other countries. We also looked for information in reports produced by the US national laboratories at Argonne, Los Alamos, and Oak Ridge. We also used reports produced by non-governmental organizations involved in nuclear nonproliferation: the Center for Nonproliferation Studies at the Monterrey Institute of International Studies and the Bellona Foundation. Finally, we searched online for articles, workshop abstracts, and websites of organizations that produce and sell the eight radioisotopes discussed here. (See **References** for complete list of sources.)

Problems

Not all or even most of the above open sources consulted provided information useful for our investigation. (It is possible that fee-based sources had more information, but these were outside the scope of this assignment.) It was relatively easy to locate nuclear facilities in the world, but it was much more difficult to determine whether or not these facilities produced radionuclides, or, if they did produce radionuclides, which if any of the eight radionuclides of particular concern were produced. Also, the information available was often contradictory. For example, a record for a facility in the IAEA database may state that the reactor is used to produce isotopes, but the Isotope Production field of that same record will show no production.

The main problem, though, is the lack of information regarding yields or facility output amounts. Even if a source clearly stated what a facility produced, it rarely stated *how much* it produced. Therefore, if we found any evidence of production of a radionuclide at a facility, we included that facility in the database.

Another problem was that many sources of information were outdated or were very biased against nuclear technologies in general. Also, not finding research reactors in countries where one would expect to see them was perplexing. For example, at the time of this submission, the continent of Africa, has 66 countries, 23 of which are not being monitored by the IAEA through formal agreements (US Department of State 2004). Of the remaining 43 countries, only two appear to have research reactor facilities producing radioisotopes, one in Egypt and one in South Africa. Twelve other countries in Africa, however, have research reactors either online, in construction, or planned (International Atomic Energy Agency 2004).

Results: Reactor Concentrations by Country

The findings of this investigation can be accessed at the SQL database located at <http://falsedrop.com/566isotopes/index.php> [password protected]. The YIELD section of the database shows that there are 36 facilities in the world that produce one or more of the eight radionuclides discussed in this paper. Not all of these facilities are research reactors. The Argentine, Canadian, and South African reactors are commercial facilities. The Russian Mayak facility is a defense complex that also produces radioisotopes for sale to civilian buyers. The Leningrad plant's primary function is to produce electricity.

Argentina

Argentina has two reactors that produce one of the isotopes being examined. The first is the Atucha 1 reactor in Cordoba and the second is the Embalse 1 reactor in Lima. Both are owned and operated by the National Commission of Atomic Energy (CNEA) of Argentina and both produce Cobalt-60 (Ferguson et al. 2003, A2-1). The CNEA reportedly controls 10 percent of the Cobalt-60 market (Tuyle et al. 2003, 35).

Australia

Australia has only one operational nuclear reactor, the HIFAR reactor located near Sydney. It is used for both commercial production and research. Only two of the radioisotopes of main concern are produced there, Cobalt-60 (Green 1997) and Iridium-192 (IAEA 2002).

Belgium

The BR-2 reactor at Mol is owned and operated by S.C.K./C.E.N., the Belgian government's research organization. This facility produces Iridium-192 for the National Institute for Radio Elements (IRE), which is also owned by the Belgian government (Ferguson et al. 2003, 29, A2-1).

Brazil

Brazil has one reactor that produces four of the isotopes being examined. This facility, IEA-R1, in Sao Paulo is owned by IPEN-CNEN/Sao Paulo and is operated by IPEN. The reactor produces Americium-241, Cesium-137, Cobalt-60 (Saxena 2001), and Iridium-192 (IAEA 2002).

Bulgaria

The IRT-SOFIA reactor in Sofia, owned by Bulgarian Academy Of Sciences and operated by Institute of Nuclear Research & Nuclear Energy, is reported to produce Cobalt-60 (Fergusen et al. 2003, A2-1).

Canada

Canada has two reactors that produce a large proportion of the world Cobalt-60 market. The first Cobalt-60 producer is the NRU reactor in Chalk River, which is owned by Atomic Energy of Canada Ltd. and operated by Chalk River Laboratories (Fergusen et al. 2003, 27). The second is the Bruce B power producer in Bruce that is owned and operated by AECL and also produces Cobalt-60 (Fergusen et al. 2003, A2-1). MDS Nordion, originally a crown corporation owned by AECL, controls 80 percent of the global Cobalt-60 market and is currently developing two new isotope production reactors Maple 1 and Maple 2 (Tuyle et al. 2003, 34).

China

China has at least four reactors that produce radioisotopes (Fergusen, et. al 2003, 32). One of these is the HFETR research reactor at Chengdu that produces Cobalt-60 and Iridium-192 (Fergusen, et. al 2004, A2-2). The Tuoli complex outside Beijing also oversees production of Californium-252, Cobalt-60, and Iridium-192 as part of a joint venture with foreign companies (Federation of American Scientists 2000).

Czech Republic

Iridium-192 is produced at the LWR-15 REZ reactor in Rez (near Prague). This facility is owned and operated by Nuclear Research Institute Rez Plc. (Fergusen et al. 2003, A2-2).

Egypt

Egypt has one Multipurpose Reactor (MPR) with a hot cell at Inshas, near Cairo, that produces sealed Cobalt-60 intended for medical use. A cyclotron accelerator was slated for commissioning in 1997 that would produce unspecified short-lived isotopes. The Egypt Atomic Energy Authority's Division of Production of Radioisotopes and Radiation Sources centers its attention around radioisotopes outside the scope of this paper (Egypt Atomic Energy Authority 2004). Nonetheless, this type of activity evidences capabilities of producing the eight radioisotopes at issue here, so continued monitoring is warranted.

Germany

The BER-II reactor in Berlin, owned by Hahn-Meitner-Institut Berlin GmbH and operated by "BER-II Reactor," is reported to produce Cobalt-60 and Iridium-192 (Green 1997). The FRJ-2 (DIDO) reactor in Juelich, owned and operated by Forschungszentrum Juelich GMBH, is reported to produce Cobalt-60 and Iridium-192 (Fergusen et al. 2003, A2-2).

India

The Reactor Operations Division of Bhabha Atomic Research Centre operates two research reactors in Trombay that produce Cobalt-60 and Iridium-192 (Fergusen et al. 2003, A2-3).

Indonesia

The GA Siwabessy MPR research reactor in Jakarta produces Iridium-192 (Fergusen et al. 2003, A2-3).

Japan

The Japan Atomic Energy Research Institute owns two research reactors that produce two of the radionuclides discussed in this paper. There is the JMTR reactor at Oarai that produces Iridium-192 and the JRR 3M reactor at Tokai that produces Iridium-192 and Cobalt-60 (Ferguson et al. 2003, A2-3).

Korea (North)

The IRT-DPRK research reactor (also known as IRT-2000 Nuclear Research Reactor) at the Yongbyon complex in the Democratic People's Republic of Korea produces several isotopes including Cobalt-60 (Center for Nonproliferation Studies 2003).

Korea (South)

The Korea Atomic Energy Research Institute owns and operates the research reactors at Yuseong (Daejeon) and at Wolsong for the production of radioisotopes. The HANARO reactor at Yuseong produces Cobalt-60 and Iridium-192 for industrial use (Korean Atomic Energy Research Institute). The Wolsong research reactor complex produces Cobalt-60 (Ferguson et al. 2003, A2-4).

Morocco

Though not yet online, the TRIGA MARK II research reactor located in the northeastern sector of Rabat is slated to develop and produce radioisotopes and radiopharmaceuticals. Unfortunately, the information on this source was last updated in mid-1999 (IAEA; The Nuclear Research Centre of Maamora).

Netherlands

The HFR reactor at Petten is owned by the European Commission and operated by the Joint Research Centre. This reactor produces radioisotopes for two major suppliers, IRE and Mallinckrodt. Ferguson et al. (2003, 30) state that HFR produces "more than 70 percent of medical radioisotopes used in European medical facilities," but the authors state that only HFR's production of Iridium-192 is a "potential, significant security risk."

Norway

The JEEP II reactor in Kjeller, owned and operated by Institutt for Energiteknikk, is reported to produce Cobalt-60 (Ferguson et al. 2003, A2-4).

Romania

The TRIGA II PITESTI - SS CORE reactor in Pitesti, owned by RAAN and operated by Institute For Nuclear Power Research, Pitesti, is reported to produce Cobalt-60 (IAEA 1998) and Iridium-192 (Ferguson et al. 2003, A2-4; IAEA 1998).

Russia

There are at least five Russian locations where the radionuclides discussed in this paper are produced. In the St. Petersburg area, there are the Leningrad Nuclear Power Plant, the Isotope Department of the Radium Institute, and the WWR-M research reactor at Gatchina. The Leningrad Nuclear Power Plant is into the radionuclide production business in addition to its main function as a producer of electricity: the plant produces Cobalt-60 (Leningrad Nuclear Power Plant, 2001). The Isotope Department of the Khlopin Radium Institute produces and sells Americium-241, Radium-226, and Strontium-90 through its commercial entity known as Ritverc GmbH (Ritverc GmbH.) The WWR-M research reactor operated by the Petersburg Nuclear Physics Institute produces Iridium-192 (Ferguson et al. 2003, A2-5).

The Center for Nonproliferation Studies at Monterrey has stated that the Scientific and Research Institute of Atomic Reactors complex in Dimitrovgrad is "of particular interest from the perspective of radiological material security" (Ferguson et al. 2003, 28). The reason for this is that the SM3 research reactor at Dimitrovgrad produces four of the eight radionuclides of concern: Americium-241, Californium-252, Cobalt-60, and Iridium-192.

A south Urals facility at Ozersk (near Chelyabinsk) is a producer of three of the eight radionuclides discussed here. This is the Mayak facility that is not a research reactor but, rather, a nuclear weapons producing facility that has been producing materials for civilian use in recent years. The two reactors here -- "Ruslan and Liudmilla" (that is, the Russian versions of "Romeo and Juliet") --

produce Cobalt-60, Iridium-192, and Plutonium-238 for domestic and foreign buyers (Center for Nonproliferation Studies 2003; Alimov 2001).

South Africa

South African Nuclear Energy Corporation is quite proud of its cutting edge nuclear technology. Its Nuclear Technology Products Division boasts that its SAFARI-1 Nuclear Research Reactor located at Pelindaba is supported by an onsite, entire production life-cycle, beginning with a supply of fuel and continuing through the storage/disposal of the resultant radioactive waste. It is an ISO 9000-certified high flux reactor, with state-of-the-art hot cells producing radioisotopes in accordance with IAEA standards and meeting ISO 9002 certification. The Corporation's web site provides information regarding its production of Cesium-137 and Iridium-192, but unfortunately amounts of production are missing (South African Nuclear Energy Corporation).

Sweden

The R2 reactor in Nykoping, owned by Studsvik AB and operated by Studsvik Nuclear AB, is reported to produce Cobalt-60 and Iridium-192 (Ferguson et al. 2003, A2-5).

Switzerland

The AGN 211 P reactor in Basel, owned by Universitaet Basel and operated by Institut Fuer Physik, Universitaet Basel, is reported to produce Cobalt-60 (Ferguson et al. 2003, A2-5).

United Kingdom

The IMPERIAL COLLEGE reactor in Ascot, owned by Imperial College of Science, Tech & Medicine and operated by the Reactor Centre, is reported to produce Cobalt-60 (Ferguson et al. 2003, A2-5).

United States

The United States has four reactors that produce the isotopes being examined. The first is the HFIR in Oak Ridge, Tennessee, which is owned by USDOE, operated by ORNL-Research Reactor Division and produces Californium-252, Iridium-192, Cobalt-60, Americium-241, Plutonium-238 (ORNL 2003). The second reactor is the NSCR Texas A reactor located in Bryan-College Station, Texas, and owned by the Texas A&M University System. NSCR Texas A is operated by Nuclear Science Center, Texas A&M University System, and produces Iridium-192. The third reactor is the WSUR Washington State University reactor in Pullman, Washington, which is owned and operated by Washington State University and produces Iridium-192. The fourth is the ATR at the INEEL site (an area 32 miles west of Idaho Falls, Idaho), which is owned by US Department of Energy and is operated by ONEEL (sometimes also operated by Isotopes International, Inc.), and produces Cobalt-60 and Iridium-192 (IAEA).

Conclusions

The authors of this paper are not qualified to make conclusions or recommendations on the disposition, safety, or security of these eight radioisotopes of most concern. However, it is acknowledged that more research into production and research reactors must be accomplished in order to gain an accurate picture of where global security risks may lie.

From the foregoing research one could conclude that of the eight radionuclides of concern, Cobalt-60 and Iridium-192 are produced by more facilities around the globe. This conclusion, however, does not take into account unreported outputs of other radionuclides. In other words, one could say that this research only shows that there is more open source information about producers of Cobalt-60 and Iridium-192 than producers of the other six radionuclides.

The best or "most honest" conclusion from this study is that there are at least 59 facilities in the world that could produce or presently produce one or more of the eight radionuclides. To state more is difficult given the sources cited here which were often less-than-definitive or exhaustive.

Recommendations for further research

Further research should determine exactly what is produced at research and commercial facilities and in what amounts in order to determine the relative risk that each individual facility poses. This determination could then be made by comparing the output amount with the levels of security given to production and distribution.

Further research could also be conducted by focusing directly on the production and presence of the radionuclides, whether that production and presence is at a reactor facility or not. For example, this research only counted facilities that are currently producing these radionuclides. It is conceivable that shut-down facilities or facilities that no longer produce one or more of the eight radioisotopes will have stockpiles of them at their facility. There are also many planned reactor facilities that will impact the production rates for the eight isotopes of most concern. Detailed research that begins with the consumers of radioisotopes could also be conducted and perhaps lead to more accurate information regarding the amounts of radioisotopes that any given facility would have on hand or be producing at a particular time. Future research via unique avenues is highly recommended.

Selected References

- Agency for Toxic Substances and Disease Registry. <<http://www.atsdr.cdc.gov>> [Accessed Mar 21 2004].
- Alimov, Rashid. 2001. *Mayak plant to increase Cobalt-60 export*. (Apr30) <http://www.bellona.no/en/international/russia/nuke_industry/siberia/mayak/20208.html> [Accessed Mar 2004, 30].
- Buxbaum, Peter A. 2004. Commercial radioactive sources seen as continuing terror threat. *Homeland Security & Defense*, 17 March. Accessed through <http://global.factiva.com> on 9 April 2004.
- Centers for Disease Control and Prevention. 2004. *CDC Radiation emergencies*. (Feb 12) <<http://www.bt.cdc.gov/radiation>> [Accessed March 2004].
- Center for Nonproliferation Studies (2003) *Russia: Mayak Production Reactors*. <http://www.nti.org/db/nisprofs/russia/fissmat/pumayak/prod_reac.htm> [Accessed Apr 12, 2004]: Nuclear Threat Initiative.
- Dictionary.com*. 2004. Lexico Publishing Group, LLC. <<http://dictionary.reference.com/search?q=Gamma%20ray>> [Accessed Apr 11, 2004].
- Egypt Atomic Energy Authority. *The Multipurpose reactor MPR*. <<http://frcu.eun/eg/www/homepage/aea/mpr.htm>> [Accessed Apr 14, 2004].
- Federation of American Scientists. 2000. *Tuoli: China Institute of Atomic Energy*. (Apr 18). <<http://www.fas.org/nuke/guide/china/facility/tuoli.htm>> [Accessed Apr 15, 2004].
- Ferguson, Charles D. 2003. *Commercial radioactive sources: Surveying the security risks*. (Mar 3). <http://www.house.gov/markey/Issues/iss_nonproliferation_st030303.pdf> [Accessed Apr 13, 2004].
- Ferguson, Charles D., Tahseen Kazi, and Judith Perera. 2003. *Commercial Radioactive Sources: Surveying the Risks*. Occasional Paper No.11. (Jan) <<http://cns.mii.edu/pubs/opapers/op11/op11.pdf>> [Accessed Apr 8, 2004].
- Green, Jim. 1997. Future Supply Of Medical Radioisotopes In Australia: Do We Need A New Reactor? <<http://www.uow.edu.au/arts/sts/pgrad/phdthesis/JimGreen/futureappend.html>> [Accessed Mar 30, 2004].
- Honour, David. 2003. *Dirty bombs—The risk and the impact*. Nov 21). <<http://www.continuitycentral.com/feature042.htm>> [Accessed Apr 11, 2004].
- International Atomic Energy Agency. *Research Reactor Data Base*. <<http://www.iaea.org/worldatom/rddb/>> [Accessed Apr 12, 2004].
- International Atomic Energy Agency. 1998. *Directory of Nuclear Research Reactors*. Vienna: IAEA.

Korean Atomic Energy Research Institute. <<http://hanaro.kaeri.re.kr:8080/english/rp.htm>> [Accessed Mar 30, 2004].

Lambert, David. Interview by the authors. Knoxville, TN. 16 March 2004.

Leningrad Nuclear Power Plant. 2001. *Proizvodstvo izotopa kobal'ta-60*. <http://www.laes.ru/new_Inpp/mindex.shtml?../content/proizv/tehnology/ort/ort.htm#02> [Accessed Apr 12, 2004].

Los Alamos National Laboratory's Chemistry Division. 2003. Periodic Table of the Elements. (Dec 11). <<http://periodic.lanl.gov/default.htm>> [Accessed Apr 14, 2004].

The Nuclear Research Centre of Maamora. <http://www.cnesten.org.ma/cnesten_eng/maamora/visite.htm> [Accessed Apr 14, 2004].

ORNL. *Index of Isotope Products*. <<http://www.ornl.gov/sci/isotopes/catalog.htm>> [Accessed Apr 14, 2004]: ORNL.

Ritverc GmbH. <<http://www.atom.nw.ru/isotope/eritverc.htm>> [Accessed Apr 2004, 12].

Saxena, R.N. *The Brazilian Research Reactor IEA-R1: Its Utilization in Basic and Applied Research and Irradiation Services*. <http://www.icens.org/Reactor_Workshop/Sessions_5__6__7/sessions_5__6__7.htm> [Accessed Apr 12, 2004]: International Centre for Environmental and Nuclear Sciences.

South African Nuclear Energy Corporation. <<http://www.pelindaba.co.za/>> [Accessed Apr 14, 2004].

Tuyle, G. J., T. L. Strub, H. A. O'Brien, C. F. Mason, and S. J. Gitomer. 2003. *Reducing RDD Concerns Related to Large Radiological Source Applications*. Los Alamos National Laboratory. (September) <http://www.nti.org/e_research/official_docs/labs/LAUR03-6%202.pdf> [Accessed Apr 15 2004].

U.S. Department of State. Country Information. *Bureau of Near Eastern Affairs and Bureau of African Affairs*. <<http://www.state.gov/p/nea/ci/>> [Accessed Apr 11, 2004].

U.S. Environmental Protection Agency. <<http://www.epa.gov>> [Accessed Mar 16, 2004].

Author Information

Mr. James Staub
 The School of Information Sciences
 University of Tennessee, Knoxville - UTK
 140 Waller Avenue
 Maryville, Kentucky 37803
 USA

jstaub@utk.edu
<http://www.utk.edu>

Grey Literature and Urban Planning: History and Accessibility

Rose M. Jackson, Portland State University

Unfortunately, Library of Congress catalog card numbers are rarely assigned to documents issued by municipal, state, and federal governments in time for inclusion in their first (and usually last) printing. In part, this is because governmental publications are considered most appropriately in the public domain, with no restrictions on quotation or reproduction; therefore, Library of Congress catalog card numbers are not assigned in advance as they frequently are for commercial publishers when copyright is requested. State and municipal governments tend to discount the vital function of libraries as depositories and central references for documentation of lasting value. Since many important documents relating to comprehensive urban planning are issued today—and will be issued increasingly in the future—by regional, state, and local governments, Library of Congress catalog card numbers or Standard Book Numbers should be available for inclusion in the initial printing. Otherwise, any electronic reference and data-processing system for city planning which utilizes general reference or standard numbers as descriptors may be seriously incomplete until the assignment of these numbers to documents is received...perhaps long after publication.¹

Melville C. Branch

Abstract

In 1970, Branch had his finger on the pulse of urban planning and its grey literature implications for urban planners. The majority of urban planning primary material, is published within the public domain with an intentionally limited distribution; therefore it rarely receives standard "cataloging" tags from the originating agency, and as a discipline, planners discount the function of a central reference point (or library) for the collection of documents. By most generally accepted definitions the data, documents and reports produced by urban planners are "born" grey. This paper examines these factors and their implications for researchers and librarians within the arena of Urban History and Planning and based upon the preliminary results of a survey offers some possible areas for collaboration.

Introduction

As a new librarian responsible for collection development within the field of Urban Studies, I first sought insight toward developing a collection philosophy by examining Urban/City Planning guides and bibliographies. Within our collection, in addition to the Branch text, I found two other guides of note: *City Planning: A Basic Bibliography of Sources and Trends* by George Bestor² and *The History of Urban Planning in the United States: An Annotated Bibliography* by Martin Gellen.³ Although these bibliographies have historical value, they lack consistency and do not adequately reflect the composition and forms of material currently being produced with the field. After contacting several librarians at universities granting a degree in Urban Studies/Planning or Architecture, I sought contact and information with the "governing" library association: the Council of Planning Librarians (CPL). Unfortunately, the CPL, an organization devoted to the collection, preservation, and exchange of planning material, disbanded in 1999. The CPL, historically, produced a directory of planning libraries. This directory facilitated the exchange of professional information and planning reports among libraries. However, the last directory was published in 1990.⁴ Therefore what was "born" grey but had some shades of "white" with respect to accessibility has moved further toward "black" among researchers and libraries. At a time when many institutions are offering a course in Urban History, the primary material supporting this field, the planning reports, documents, pamphlets, and drawings, are being lost.

¹ Branch, Melville C. *Comprehensive Urban Planning: A Selective Annotated Bibliography with Related Materials*. Beverly Hills: Sage, 1970.

² Bestor, George C., and Holway R. Jones. *City Planning: A Basic Bibliography of Sources and Trends*. North Sacramento: The Central Printing Company, 1962.

³ Gellen, Martin. "The History of Urban Planning in the United States: An Annotated Bibliography." CPL Bibliographies 107(March 1983).

⁴ Council of Planning Librarians. *Directory of Planning and Urban Affairs Libraries in the United States and Canada*. 5th Ed. Ed. Patricia Coatsworth. Chicago: ALA, 1990.

History

Though planning as an aspect of architecture has been in existence for thousands of years, the field of Urban Studies; the codification of comprehensive planning within a metropolitan area: zoning, transportation, citizen surveys, project reports, the combination of politics, engineering, and art, has more recent origins. The birth of urban planning within the United States parallels that of public and private libraries. Public and private firms engaged in planning maintained in-house libraries for the collection, dissemination, and preservation of key documents, reports and literature. Between the 1960s and 1990s, there was a substantial enough investment in these libraries and the material being produced that urban planning professionals and librarians felt the need to create bibliographic guides and associations to support the developing field.

However, at the dawn of the 21st century, the efforts aimed at collection development within urban planning are taking place in silos and in isolation. With the advent of new technologies such as Geographical Information Systems (GIS), which allows for the longitudinal as well as the geographical overlay of data, the need for new standards and transfer of information requires some overarching organization and uniformity. Indeed, the influential leaders who created CPL and participated in these early organizations could not have predicted the rapid acceptance and implementation of new technologies that would allow the collection to move from a physical to a virtual state, accessible 24 hours a day. Lacking a national or international organization with the mission to embrace these new technologies and develop a consistent set of standards for their implementation has resulted in what was once moving toward "white" in terms of accessibility now regressing toward "black."

In the Public Domain

Urban Studies is a unique discipline in that a significant amount of the primary literature is produced by non-profit or government agencies that do not publish or actively promote the sale of their literature. Most municipal agencies, as part of their mission, mandate that most of their reports, documents, and data be available to the general public. Though Branch was correct in 1970 in stating that this material is within "the public domain," copyright restrictions have changed. Governmental agencies do subcontract research. Additionally, some municipal agencies will charge individuals and private firms for the creation and maintenance of agency reports. These contracted reports have certain restrictions with respect to public access and reproduction. However, most urban planners work in governmental (public) firms. Yet, their philosophical approach to collection development and maintenance adheres more toward an architectural (private firm) approach. Even though, the majority of their documents are not produced for profit - instead of adopting the model of the Federal Depository Library Plan- most state and municipal agencies emulate for-profit firms and treat internally produced documents and reports as proprietary.

Not only do the agencies' treat this data as proprietary to their agency but also within the agencies, there is a distinct lack of sharing amongst planning teams. If a researcher knows or has some certainty that a particular report exists, the burden is upon the researcher to go through an iterative dance with the agency and specific planners within the firm to obtain a copy of the report. This lack of transparency with respect to the production and location of data requires urban planning librarians to spend an inordinate amount of time developing strategic searching skills and "information network" and then teaching new scholars and researchers to identify and access the desired material. At every firm and academic institution, the reference team works to develop a systematic approach to addressing this issue. What could be viewed as an advantage for collection development, limited copyright restrictions, cannot by itself overcome other barriers to collection.

At the same time, these librarians must seek out sources of grey literature within the field and work within their own organization to "capture" it. Since there are few copyright restrictions, the major limiters on adding these documents to the collection are locating the material and then performing original cataloging.

The cost of original cataloging could exceed \$50.00 per item depending on the level of complexity and format of the material.⁵ Original cataloging requires not only a professional skilled in the rules of cataloging but a subject specialist with in-depth knowledge of the field. Such talent is rare, which results in increased costs for the institution employing or attempting to contract with them.

Standardized Cataloging Schema

Another contributing factor affecting accessibility is the lack of agreement concerning library collection practices with respect to urban planning material. The material is produced within the public domain and initially is viewed as having limited value. However, an emerging trend is that the value of these documents increases with time as researchers within the field and in other fields attempt to frame comprehensive planning within the context of the prevailing political, social, and economic climate.⁶ In

⁵ FEDLINK Technical Processing Services <http://www.loc.gov/flicc/prices/tppr-ori.html> retrieved November 5, 2004.

⁶ Osif, Bonnie A. "Transportation Information: A Review of Grey Literature by Format, Language, and Availability." *The International Journal on Grey Literature*. 1(2000): 12-17.

the 21st century, those internally produced, non-cataloged documents with limited distribution have additional currency and meaning beyond their initial purposes.

Branch's second point, the lack of standardized cataloging, is a key stumbling block to accessing this material. Since the majority of these resources were produced for limited distribution, the initial cataloging did not necessarily adhere to a national or international schema. Historically, most in-house librarians developed their own schema for indexing the material. The material could still be retrieved either by the author, date of publication, subject field, or title (the assumption being that the librarian has an intimate knowledge of the collection). This lack of standardization in cataloging could impact the collection's conversion/inclusion in online databases/catalogs. For a library to conduct a retrospective conversion of its older print collection to a current cataloging standard such as Anglo-American Cataloging Rules, 2nd edition (AACR2) and include these bibliographic records in its current catalog could take several years. Hampering this conversion is the fact that solo practitioners maintain most of these in-house libraries. If state and municipal agencies sought a Library of Congress CIP or an ISBN for works they planned to publish (as Branch suggested), it then would make it easier for other libraries to be aware of the publication of material and collect it. Again, this assumes the originating organization produced enough copies in the print run. However, the effort to obtain this type of documentation exceeds the staff constraints of many agencies. How then is a researcher to obtain the material?

Central Depository

Branch's last lament would be the modern researcher's first option; seek out a central depository or library that focuses on Urban Planning. A central repository would negate the need to increase the print run to satisfy demand for access.

There are over 100 universities and colleges within the United States that offer a degree within the field of Urban Studies. A researcher focusing upon a major urban development initiative within the Portland, Oregon Metropolitan Area would anticipate that Portland State University with a College of Urban and Public Affairs would house copies of pertinent documents and data related to this endeavor. This would be a false assumption. University libraries' grey literature collections and their collection development policies have lagged behind that of standard monographs and serials. Julia Gelfand's description from 1997 is still apt:

The dilemma of conflicting priorities and the controversies inherent over stretching budgets, when libraries, consider the special needs required by grey literature, become compounded. These needs may include: developing human resources with special expertise and subject background to solicit and become aware of the literature; requiring original cataloguing and special processing to make it easily, expediently and readily available; and, perhaps, creating storage systems offering protection for long-term retention.⁷

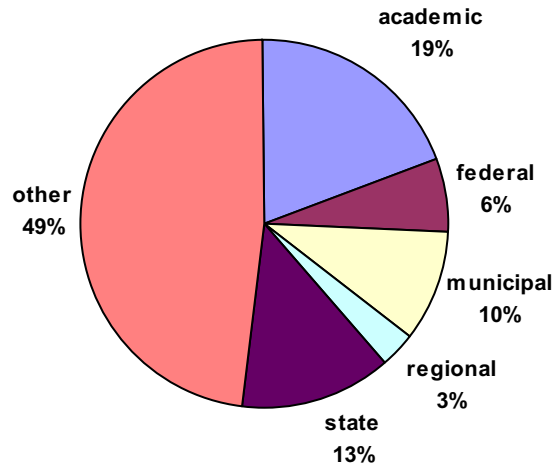
Given the chaos within academic libraries regarding the value of central depositories within the field of Urban Studies, what can researchers and librarians build upon?

Determining a Baseline

In August 2004, after receiving institutional review approval, I mailed a postcard to the 261 libraries listed in the Directory of Planning and Urban Affairs Libraries in the United States and Canada 1990. The postcard was an invitation to participate in a survey addressing four issues: library holdings, classification/cataloging schema, public accessibility, and content of their grey literature collection. In September, I posted the text of the postcard on several national listservs and emailed the text and survey URL directly to several institutions. As of October 28, 2004, I received 32 responses. Though this was a limited response, the data yields some interesting baseline information with respect to both quantitative and qualitative factors. The survey questionnaire is included as an appendix to this article. Verifying that no two individuals from the same institution responded, I was able to identify among the participants a wide range of library types: academic, federal, municipal, regional, and state. Nearly half of the responding libraries described themselves as "other": special, private, nonprofit, or a transportation library.

⁷ Gelfand, Julia M. "Academic Libraries and Collection Development Implications for Grey Literature." *Publishing Research Quarterly* 13 (Summer 1997): 15-23.

Library Type

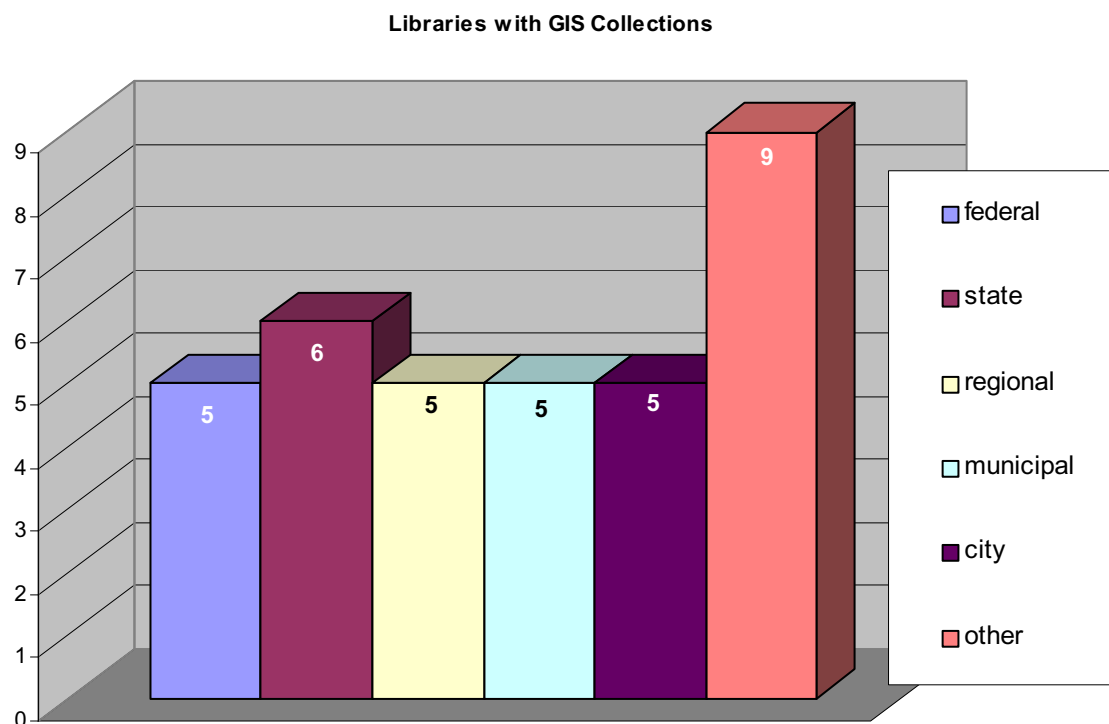


Except for those associated with universities, these libraries tended (on average) to have collections composed of:

1. ~15,000 monographs,
2. < 200 print serials,
3. < 30 electronic serials,
4. zero to very limited DVD and digital data collections, and
5. 15-5,000 maps.

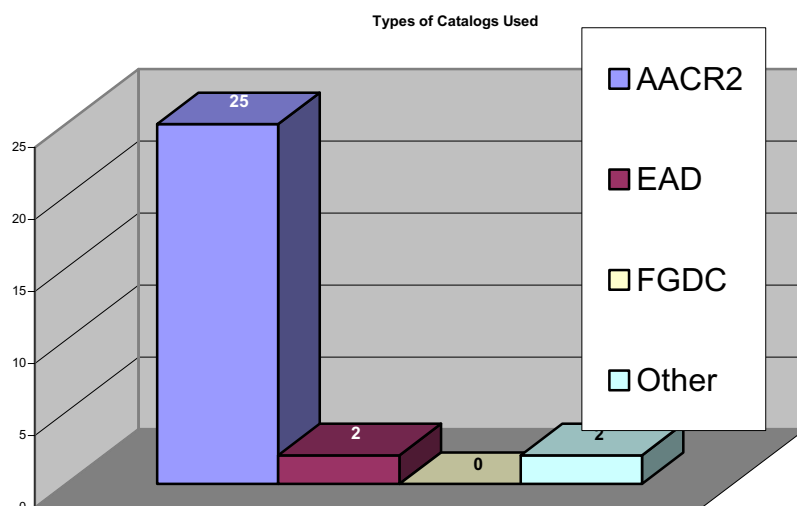
Those associated with either research universities or public libraries had significantly larger collections with greater variability. Most of these respondents reported:

1. ~ 45,000-250,000 monographs
2. >200 to 8,000 print serials
3. 350-20,000 electronic serials
4. again, zero to very limited DVD and digital data collections except for one institution, and
5. zero to 6500 maps.



While most of the libraries report some level of GIS collection, the most significant comment concerning this area of collection development was that this is a growing field or a split responsibility within the parent organization. Several of the libraries report that they have access to the GIS data and documents but do not or are not responsible for maintaining it. Among the respondents, those answering in the affirmative for having a GIS collection tended to be special libraries. Among academic institutions, the GIS material is more likely to be stored with other cartographic materials in a different branch or location. Some reporting libraries collect GIS material from more than one type of agency.

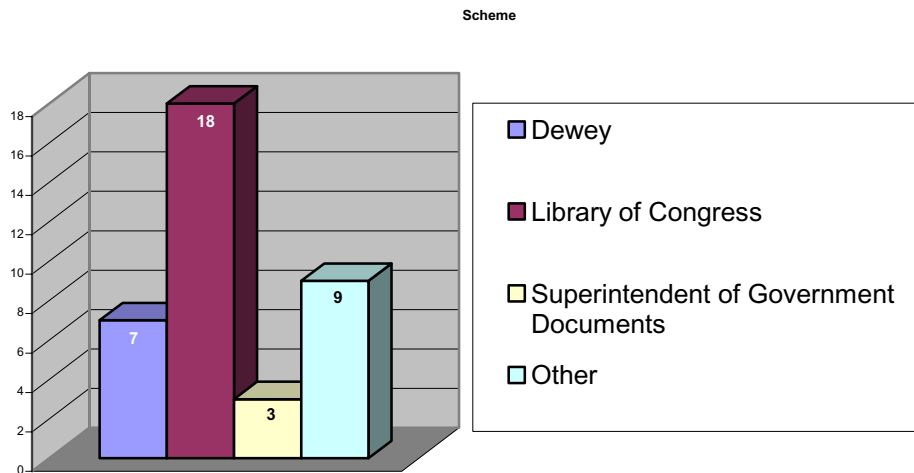
Cataloging Type:



Interestingly, most of the responding libraries are using AACR2 rules for cataloging their materials. Of the "other" respondents, one library lacked any cataloging system and the other used an in-house (homegrown) cataloging system. With the development of geo-spatial metadata standards in 1998 and

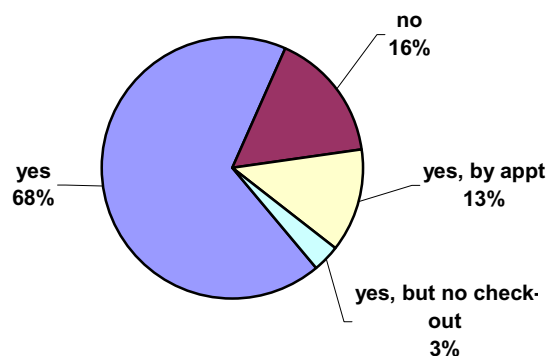
the growth in the production and use of GIS, I had assumed that some libraries would report using FGDC, particularly if they had a large GIS collection.⁸

Classification Scheme:



Most responding libraries are using a standardized scheme that would facilitate the sharing of information. However, among the nine libraries responding "other," there was no uniformity. Each employed a different schema for classification ranging from a modified Dewey to the universal decimal system. Therefore, based on the comments from the respondents, most of these planning libraries are, with respect to part of the collection, utilizing a consistent, national cataloging standard and classification schema for their libraries. Lack of consistency may not be a stumbling block to the exchange and access of data, documents, and reports maintained by planning libraries.

Libraries Open to the Public



Eighty-four percent of the responding libraries are open to the public. More reported being open to researchers within the field.

Within the qualitative section of the survey, I attempted to elicit information on the "state" of the collection, what percentage of it had been produced prior to the 1990 survey conducted by the CPL, what was its preservation status, how much had been weeded, is there a collection retention policy in effect, and whether or not its technological format made it inaccessible. Because the responses were supplied in unlimited text message boxes, the results listed below are just a rough analysis of the information provided.

⁸ McGlamery, Patrick. "Lifting the Fog: Maps and Spatial Information Emerging as Grey Literature." *The International Journal on Grey Literature*. 1(2000): 5-11.

Only two libraries were unable to estimate what percentage of their collection predates 1990.

< 20%	> 20% < 40%	>40% <60%	>60% <80%	>80%
2	7	7	5	5

Given that such a high percentage of the collection is older than 1990, it was interesting to discover that most of it is still within the collection. Almost 75% of those surveyed reported that they had no plans to withdraw the material due to either space or lack of use. My favorite response was:

So far it has been nothing short of a miracle as the collection has sustained 2 major building fires in 50 years. The library is in the process of being rebuilt from our last MAJOR building fire with a proposed completion date of January 2005. The collection is in an environmentally controlled warehouse in another county at the present time and cannot be accessed by either library personnel {or} departmental staff.

Some libraries report that older material has been removed from the circulating collection for preservation reasons.

Given the high rate of retention, it was surprising to find that less than half of the survey participants have a written retention policy. Two libraries did report, however, that Homeland Security issues post 9/11 have impacted on their implementation of retention policies. Though most libraries lack a written retention policy, they have an implicit one in effect: everything subject for removal is first reviewed by the subject specialist. One library reports: "A new emphasis has been placed on retaining older materials now that a course on the history of planning is now being taught."

When asked: "What material, produced prior to 1990, is the library retaining and why? Please give some examples," most libraries provided some very insightful and strategic answers that recognize the importance of the collection and its value to the community and the broader society. Examples include:

Every "final" report that Leigh Fisher Associates produces is stored in two places—two copies in offsite storage and one copy in-house. Ideally, we are trying to retain at least three copies of each "final"/deliverable report going back to 1946 when the company was founded. We also are retaining material that relates to "landmark" projects, for example, our early work in Tampa, and in El Paso.

For example, I have retained file materials and reports on old urban renewal projects, particularly on property acquisition, because there are often recurring issues with these properties, and how they were purchased. I also keep key documents on things like the Open Space grants, since as these areas are reused for new purposes, there are always questions about how the use of open space grant areas was restricted and can it be used for a different purpose.

The preservation of the collections varies from "not very good" to "we are digitizing collection to preserve it." From the responses, it is obvious that preservation is a significant issue in these libraries. Some have developed formal preservation processes and procedures, including climate control, while others report no special efforts have been made.

Approximately two-thirds of the libraries reported some concerns with incompatibility of older technologies. Survey participants report having 5 ¼" floppies, older databases in obsolete software packages, beta videotapes, and the well-known consistent problems with microfiche/card/film readers.

Interestingly, a significant number of the reporting libraries have an online public access catalog system in place and report that their urban planning material are cataloged within it. Some libraries report that the older material may be in either a Microsoft Access database, specialist prepared bibliographies in print, or professionally generated paper indexes. One library reports that they have a separate catalog for their older planning documents and about 25 per week are being cataloged for inclusion in their online system.

Perhaps some of the most interesting comments received pertain to the question: "Would you be interested in reestablishing the Council of Planning Librarians? Why or why not?" The results fell into 3 major categories. The first was a straight no, due to lack of time or cost concerns. The second most frequent response was a qualified "maybe" because they either knew nothing about the organization or felt the need to participate in a larger organization to communicate and compare notes. The third group replied "yes" with comments such as: "The council would provide a sense of continuity to new librarians entering the field" and "[I]t is becoming impossible to find who has what, much less acquire copies."

Yet when asked in the next question, "[w]hether their library would be interested in establishing a formal interlibrary loan agreement amongst Urban Planning libraries?" the first responder quoted above replied "no." The responses were pretty evenly distributed among "yes" and "no" votes. Some libraries based on policy are unwilling to lend material, their primary purpose being to support their local users. Others felt they could make some arrangements to lend material, particularly if it would result in lower borrowing costs for their institutions, and still others felt that participation would require a higher level of management approval and the required paperwork might take too long.

Conclusion

Branch was succinct in highlighting three of the major facets (limited distribution, lack of standards, and lack of a central repository), affecting the status of resources within the field. In addition to his points, the lack of a centralized library organization or agency to promote and develop standards, and lobby for access to this vital information has hampered urban planning research within the United States.

At Portland State University we find ourselves in the a state similar to that of the majority of respondents. We have a nationally recognized Urban Studies program, and several municipal planning departments that are constantly referenced in presentations and publications. We have cutting-edge urban planning that impacts the local, regional, state, national and international arena. Yet, we lack a central depository for either a print or digital collection of material for our students, faculty, researchers and citizen scholars. This year, with the support of the College of Urban Studies, the library applied for a Library Services and Technology Grant to create a Sustainable Digital Library to collect material from Metro, our regional government that controls the Urban Growth Boundary, TriMet, our regional transportation system, and solicit material from the county and city planners to support urban planning research on the Pacific Northwest.

Our goal is to create a pilot project that other universities and regions with significant planning collections could access and replicate similar to the Transportation Library Connectivity Project.⁹ We hope through this endeavor to promote, encourage, and facilitate the formulation of a "networked" organization that can fill the void left after the demise of the Council of Planning Librarians. Based on their hours of operation this tacit recognition of need highlights the respondents' commitment to making the material available to interested parties. These libraries and librarians recognize that they are working in a vacuum concerning professional support and organization but are uncertain about participating in a national/international organization and whether or not their institution would benefit from participation in a formal interlibrary loan arrangement.

These print collections contain a significant amount of valuable, archival material that is being used on a regular basis by patrons. Documents, data, and reports produced prior to 1990 comprise more than half of the collection within these libraries. The preservation and maintenance of the collection is important to the organization and retention policies are in place. Policies need to be developed for archiving software dependent data, like GIS. Now we need to take the next step together to make this material accessible to all.

Author Information

Ms. Rose M. Jackson
Reference Librarian
Portland State University
Portland, OR 97207
USA

rosej@pdx.edu
<http://www.lib.pdx.edu>

⁹ Transportation Library Connectivity. <http://www.pooledfund.org/documents/solicitations/910.pdf> retrieved November 5, 2004.

Appendices

List of Participating Organizations	178-179
GreyNet Subscription Form	180
Index to Authors	181
GL6 Publication Order Form	182

List of Participating Organizations

Aberdeen Medical School; Department of Public Health	United Kingdom
Albert R. Mann Library	USA
American Psychological Association, APA	USA
Boer & Croon, Strategy and Management Group	Netherlands
Branford P. Millar Library	USA
British Library, BL	United Kingdom
CCLRC – Rutherford Appleton Laboratory	United Kingdom
Center for Natural Language Processing, CNLP	USA
Central Liverpool Primary Care Trust, PCT	United Kingdom
Centre National de Recherche Scientifique, CNRS	France
Columbia University	USA
Cornell University	USA
CYBEK	USA
Dalhousie University; Faculty of Management	Canada
Ecole des Mines de Nantes, EMN	France
Ehrman Medical Library	USA
Endeavor Information Systems	USA
Electronic Publishing Initiative at Columbia, EPIC	USA
Emerald	United Kingdom
Grey Literature Network Service, GreyNet	Netherlands
Hamilton Fish Institute on School and Community Violence	USA
Information International Associates, Inc. IIa	USA
Institut de l'Information Scientifique et Technique, INIST	France
Institute of Information Science and Technologies, ISTI-CNR	Italy
Istituto di Ricerche sulla Popolazione e le Politiche Sociali, IRPPS	Italy
International Atomic Energy Agency, IAEA	Austria
Japan Science and Technology Agency, JST	Japan
J.W. Scott Health Sciences Library	Canada
Library of Zeeland	Netherlands
LINK s.r.l.	Italy
Long Island University; Palmer School	USA
NASA Goddard Space Flight Center Library	USA

National Diet Library, NDL	Japan
National Institute of Occupational Safety and Prevention, ISPESL	Italy
National Library of Medicine, NLM	USA
Netherlands Institute for Scientific Information Services, NIWI	Netherlands
New York Academy of Medicine, NYAM	USA
New York Public Library	USA
New York University, NYU	USA
New York University School of Medicine	USA
Office of Scientific and Technical Information, OSTI	USA
Oregon Health Science University	USA
Oregon State University Libraries; Hatfield Marine Science Center	USA
Portland State University, College of Urban and Public Affairs	USA
Pratt Institute, School of Information and Library Science	USA
Research Libraries Group, RLG	USA
Rotch Library, MIT Libraries	USA
Royal Netherlands Academy of Arts and Sciences, KNAW	Netherlands
Scientific and Technical Information Centre of Russia, VNTIC	Russia
Southeastern Louisiana University, SELU	USA
Syracuse University; School in Information Studies	USA
TextRelease - Professional Information Services	Netherlands
United States Department of Energy, DOE	USA
University of Aberdeen	United Kingdom
University of Alberta	Canada
University of Amsterdam, UvA	Netherlands
University of Bergen, UiB	Norway
University of California, Irvine Libraries, UCI	USA
University of Houston, UH	USA
University of Pretoria, UP	South Africa
University of Tennessee, Knoxville; School of Information Sciences, UTK	USA
University of Twente	Netherlands
University of Valencia, Social Science Library	Spain
University of Virginia, Health Sciences Library	USA
University of Washington Information School	USA



Grey Literature Network Service

Individual Membership 2006|07

Twelve month, all-in membership, annually renewable for only **€ 245**

"Your GreyNet Membership includes information products, services, and considerable savings"

Annual printed
edition of the
GL Conference
Proceedings
ISSN 1386-2316

Monthly printed
edition of the
GL Conference
Memoranda
ISSN 1574-1214

Enrolment in
GreyNet's
moderated
Listserv
listserv@greynet.org

Full-text
Document Delivery
via
GreyText Archive
info@greynet.org

Your
Organization's
listing in the
GreySource Index
www.greynet.org

 20%
Reduction
4-5 December 2006
GL-Conference Series

 20%
Reduction
The Grey Journal
journal@greynet.org

GreyNet's
Annual Award Dinner
Personal Invitation
December 3, 2006

PLUS: A guarantee that future GreyNet products and services will include benefits for Members!

Member information:

Name:

Organization:

Postal Address:

City/Code/Country:

Tel/Fax/Email:

Method of Payment:

- ☐ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam, Netherlands
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to "GreyNet Membership"
- ☐ Bank check/draft made payable to TextRelease, with reference to "GreyNet Membership"
- ☐ MasterCard/Eurocard ☐ Visa card ☐ American Express

Card No. _____ Expiration Date: _____

Print the name that appears on the credit card, here _____

Signature: _____ **CVC II code:** _____ (Last 3 digits on signature side of card)

Place: _____ Date: _____

NOTE On receipt of payment, an invoice marked paid will be sent to your postal address. If you require a *pro forma* invoice, please notify us!

GreyNet is powered by TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands □ T/F +31-(0) 20-331.2420
www.greynet.org www.textrelease.com

Index to Authors

Anderson, N.	24	Keith, L.	123
Asserson, A.	84	Knight, J.	91
		Krichel, T.	35
Barrueco, J.M.	35		
Biagioni, S.	41	LaFleur, L.J.	30
Blaaij, C. de	106	Lehman, H.	136
Boekhorst, A.K.	1	Liddy, E.D.	46
Bourasseau, C.	68	Luzi, D.	94
Bradley, J.	46		
		Manco, M.	94
Carman, J.	160	Mandava, L.	91
Castriotta, M.	94	Mazzucco, R.	94
Cooke, M.	79	Milewski, S.	160
Corieri, S.B.	46	Myohanen, L.	123
Costigan, S.	79		
Creel, A.	160	Paillassard, P.	143
Cutler, D.E.	64	Pavlov, L.P.	113
Di Cesare, R.	41	Ranger, S.L.	128
Dijk, E.M.S.	16	Reid, G.	91
Dumas, C.	68	Ruggieri, R.	41
Dunn, K.	50	Rupp, N.	30
Farace, D.J.	1, 55	Schöpfel, J.	55, 143
Frantzen, J.	1, 55	Simkhada, P.	91
		Staub, J.	160
Gelfand, J.	10	Stock, C.	55, 143
Giannini, S.	41		
		Taylor, E.	123
Hofe, H. von	119	Turner, A.M.	46
Jackson, R.M.	169	Webster, J.	136
Japzon, A.C.	24	Wheatley, J.A.	46
Jeffery, K.G.	84	Woofter, A.	160
		Wyness, L.	91



Publication Order Form

Conference Proceedings on Grey Literature in cooperation with *GreyNet*, Grey Literature Network Service

Publications Now in Print:	No. of Copies	x	Amount in Euros	Subtotal
GL7 CONFERENCE PROCEEDINGS - Printed Edition ISBN 90-77484-06-X <i>Postage and Handling excluded*</i>		x	70.00 =	€ (INCL. VAT/BTW)
GL6 CONFERENCE PROCEEDINGS - Printed Edition ISBN 90-77484-04-3 <i>Postage and Handling excluded*</i>		x	70.00 =	€ (INCL. VAT/BTW)
GL5 CONFERENCE PROCEEDINGS - Printed Edition ISBN 90-77484-01-9 <i>Postage and Handling excluded*</i>		x	60.00 =	€ (INCL. VAT/BTW)
POSTAGE AND HANDLING PER PRINTED COPY*				
	Holland		x 5.00 =	€
	Europe		x 10.00 =	€
	Other		x 20.00 =	€
TOTAL AMOUNT:				€ (INCL. VAT/BTW)

Customer information

Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
Fax/Email Address:	

Check one of the boxes below for your Method of Payment:

- ☐ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
BIC: RABONL2U IBAN: NL70 RABO 0313 5853 42, with reference to "Publication Order"
- ☐ Bank check/draft made payable to TextRelease, with reference to "Publication Order"
- ☐ MasterCard/Eurocard ☐ Visa card ☐ American Express

Card No. _____ Expiration Date: _____

Print the name that appears on the credit card, here _____

Signature: _____ CVC II code: _____ (Last 3 digits on signature side of card)

Place: _____ Date: _____

NOTE On receipt of payment, an invoice marked paid will be sent to your postal address. If you require a *pro forma* invoice, please notify:

TextRelease
www.textrelease.com

GL Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: info@textrelease.com