

An International Journal on Grey Literature



Spring 2010 – TGJ Volume 6, Number 1

‘GOVERNMENT ALLIANCE TO GREY LITERATURE’

GreyNet

The Grey Journal

An International Journal on Grey Literature

COLOPHON

Journal Editor:

Dr. Dominic J. Farace
Grey Literature Network Service
GreyNet, The Netherlands
journal@greyjournal.org
<http://www.greyjournal.org/thegreyjournal.html>

Associate Editors:

Julia Gelfand
University of California, Irvine
UCI, United States

Dr. Joachim Schöpfel
Université Charles de Gaulle Lille 3
France

Gretta E. Siegel
Portland State University
PSU, United States

Kate Wittenberg
Electronic Publishing Initiative at Columbia, EPIC
Columbia University, United States

Technical Editor:

Jerry Frantzen, TextRelease

CIP

The Grey Journal (TGJ): An international journal on grey literature / D.J. Farace (Journal Editor); J. Frantzen (Technical Editor) ; GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 6, Number 1, spring 2010. - *The British Library, INIST-CNRS, NYAM, and NTK are corporate authors and associate members of GreyNet. This serial publication appears three times a year - in spring, summer, and autumn. Each issue is thematic and deals with one or more related topics in the field of grey literature. The Grey Journal appears both in print and electronic formats.* ISSN 1574-1796 (Print)
ISSN 1574-180X (E-Print/PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:

€125 individual, including postage & handling
€240 institutional, including postage & handling

Contact Address:

Reprint Service, Back Issues, Document Delivery,
Advertising, Inserts, Subscriptions:

TextRelease
Javastraat 194-HS
1095 CP Amsterdam
Netherlands
T/F +31 (0) 20 331.2420
info@textrelease.com
<http://www.textrelease.com/gjpublications.html>

About TGJ

The Grey Journal is a flagship journal for the grey literature community. It crosses continents, disciplines, and sectors both public and private. The Grey Journal not only deals with the topic of grey literature but is itself a document type classified as grey literature. It is akin to other grey serial publications, such as conference proceedings, reports, working papers, etc.



The Grey Journal is geared to Colleges and Schools of Library and Information Studies, as well as, information professionals, who produce, publish, process, manage, disseminate, and use grey literature e.g. researchers, editors, librarians, documentalists, archivists, journalists, intermediaries, etc.

About GreyNet

The Grey Literature Network Services was established in order to facilitate dialog, research, and communication between persons and organizations in the field of grey literature. GreyNet further seeks to identify and distribute information on and about grey literature in networked environments. Its main activities include the International Conference Series on Grey Literature, the creation and maintenance of web-based resources, a moderated Listserv, and The Grey Journal. GreyNet is also engaged in the development of distance learning courses for graduate and post-graduate students, as well as workshops and seminars for practitioners.

Full-Text License Agreement

In 2004, TextRelease entered into an electronic licensing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The full text of articles in The Grey Journal (TGJ) can be found in *Library, Information Science & Technology Abstracts (LISTA)* full-text database.

© 2010 TextRelease

Copyright, all rights reserved. No part of this publication may be reproduced, stored in or introduced into a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior permission of the publisher.

Contents

"GOVERNMENT ALLIANCE TO GREY LITERATURE"

A Multi-Institutional Approach to Technical Report Literature: Development of the Technical Report Archive & Image Library (TRAIL)	5
Maliaca Oxnam (<i>United States</i>)	
Digitizing Grey Literature from the <i>Antarctic Bibliography</i> Collection	9
Tina Gheen and Sue Olmsted (<i>United States</i>)	
Web 2.0 Utilization in E-Diplomacy and the Proliferation of Government Grey Literature	15
Jonathan Omowalé Cain (<i>United States</i>)	
The Grey System for Monitoring Self-Funded Research	25
Aleksandr V. Starovoitov, Yuri M. Bogdanov, Aleksandr M. Bastrykin, and Leonid P. Pavlov (<i>Russia</i>)	
An Infrastructure for Managing EC Funded Research Output: The OpenAIRE Project	31
Paolo Manghi (<i>Italy</i>), Natalia Manola (<i>Greece</i>), Wolfram Horstmann (<i>Germany</i>), and Dale Peters (<i>South Africa</i>)	
OpenSIGLE - Crossroads for Libraries, Research and Educational Institutions in the field of Grey Literature	41
Dominic J. Farace and Jerry Frantzen (<i>Netherlands</i>), Christiane Stock, Nathalie Henrot, and Joachim Schöpfel (<i>France</i>)	

Colophon.....	2
Editor's Note.....	4
On the News Front	
GL12 Call for Papers, Twelfth International Conference on Grey Literature.....	49
Transparency in Grey Literature, Grey Tech Approaches to High Tech Issues	
Advertisements	
NTK – Czech National Technical Library	8
Ila – Information International Associates, Inc.....	24
J-STAGE – Japan Science and Technology Agency.....	30
Refdoc.fr – Institute for Scientific and Technical Information.....	40
About the Authors.....	50
Notes for Contributors.....	51

EDITOR'S NOTE

At the onset of Twenty-ten, Grey Literature emerged into the public arena after more than a quarter century in the corridors of libraries and in workplaces and meeting rooms of information practitioners and professionals. Grey Literature is now a topic of news in the world media. Coverage in magazines and newspapers e.g. Nature, New Scientist, The Economist, the Guardian, etc. carrying articles on the IPCC use/misuse of grey literature is current and in-depth. For those following these news threads, much of the publicity is less than complimentary. And, the grey literature community has not been hesitant in their response via blogs, listservs, distribution lists, etcetera. During the coming months leading up to GL12, the international grey literature community will have the opportunity to bundle its efforts in order to address issues that stand at the core of grey literature and which have come under fire in the public media. One thing is certain, now that grey literature has entered the mainstream press, it will not simply disappear. It is now up to the corporate authors and publishers of grey literature as well as those organizations processing and distributing it both in print and electronic formats to address the misconceptions and unknowns about this field of information science. The Twelfth International Conference on Grey Literature will provide a global forum for stakeholders in government, academics, business and industry to come together on issues formulated in the GL12 Call-for-Papers. This year's proposed themes accentuate the transparency in grey literature and the almost seamless processes of research, authorship, publication, indexing, as well as, the uses and applications to which it is exposed in knowledge based communities. Many of these processes are the same faced by commercial publishing, where only the differences lie in grey tech approaches to high tech issues.

Dominic J. Farace, Journal Editor
journal@grey.net.org

Grey Literature "information produced and distributed on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body." (Luxembourg, 1997 expanded in New York, 2004)

A Multi-Institutional Approach to Technical Report Literature: Development of the Technical Report Archive & Image Library (TRAIL)*

Maliaca Oxnam
(United States)

Abstract

While availability and access to more recent technical report literature has greatly improved through electronic delivery, older technical report literature still remains elusive to users. The challenges posed by these collections are two-fold: 1) the diverse nature of distribution and collection building practices across institutions; and, 2) the variability in the selection and acquisition methods used, resulting in incomplete series, confusion in depository and non-depository status, lack of awareness and definition of a "complete collection," and a host of other inconsistencies.

Through a national dialogue among academic librarians led by University of Arizona Libraries (UAL) beginning in 2005, common discoveries were made about the various barriers libraries were experiencing in managing and making accessible legacy technical report collections, and the urgent need to address these issues. The positive response to these dialogues and preliminary assessment results led to an extraordinary opportunity to resolve a difficult challenge in academic and special libraries. These problems created a strong case for establishing the Technical Report Archive & Image Library (TRAIL) collectively maintained by the library community.

The Technical Report Archive & Image Library (TRAIL) aims to identify, digitize, archive, and provide persistent and unrestricted access to federal technical reports issued prior to 1975. TRAIL is a Greater Western Library Alliance initiative led by the University of Arizona in collaboration with the Center for Research Libraries. TRAIL is funded by the member institutions of the Greater Western Library Alliance with an estimated annual budget of \$128,000. To date, TRAIL has scanned more than one million pages of technical reports issued by the US federal government agencies and has involved participation and contribution from more than twenty universities. This paper will describe TRAIL, including the need for its development, current structure, and future directions.

Why a focus on federally funded technical reports?

In 2004, the University of Arizona Libraries (UAL) focused efforts on eliminating backlogs of purchased materials needing to be processed and cataloged. One of the results of this focused effort was the recognition that our library had collections of materials, sitting on the shelves available to users, that were deemed valuable, but with little or no title level cataloging, or other finding aids available. These collections were virtually "hidden" from our users. While available, most of these collections required that users employ some form of mediation or assistance from the library, which was the exact opposite direction from the Library's strategic direction at the time for creating "self-sufficient users" who could access the library "anytime, anyplace". One of the identified collections was a large collection of print technical research report series from various US federal agencies. The collection was housed in our Science-Engineering Library, separate from our other government document collections. These collections were purchased prior to UAL's establishment as a selective depository library and/or were purchased directly from the agencies and hence were not part of our official government documents depository collection. In addition to our print collection, there were also many thousands of reports that we owned issued either in micro-opaque or micro-card format and microfiche format. To further the complexity, these collections were full of gaps in the numbering sequences of the report series. These gaps were caused by reports that weren't publicly issued, reports that were not available at the time of the purchase of our collections and never filled in as the reports may later have been released, and also likely to natural theft and loss over time.

* First published in the GL11 Conference Proceedings, March 2010

The UAL spent considerable time trying to locate shipping and/or receiving lists to get an item level inventory for all the technical reports we'd received in our collection. It was hoped that we might be able to create some form of inventory of items that could be added to our local OPAC. However, it didn't take us long to realize that we weren't going to be able to find records of what items had been shipped and/or received, and hiring students to manually inventory our collection was going to be time-consuming and costly. Providing better access to these collections wasn't a project that could realistically be undertaken by any one institution. And UAL temporarily ceased its project until funding and/or technologies to help with the project could be secured.

In the summer of 2005, a discussion on technical report collection access was brought forward to interested members of the Engineering Libraries Division of the American Society of Engineering Education. A group email list was started for the more than 80 individuals who expressed interest in creating better access to these early technical report collections that so many of us held in our individual library collections. Ideas for preferred access points for libraries and users were discussed. Unanimously, librarians expressed the desire to have online access to these older reports with the recognition that few of the federal agencies were budgeted to digitize or would be able to provide online access to all of the older materials held in our collections. Through discussion, it was reinforced that many libraries, including government depository libraries and federal agency libraries held incomplete collections. To fulfill user requests for items, we really needed the collective whole of all the collections in order to ensure that any given user request could be met. Discussions also led to informal surveys regarding which report collections were thought to be most useful and most desired by the group as a whole.

Development of the Technical Report Archive & Image Library

Later the next year, after discussing the idea of a collaborative project with the Center for Research Libraries (CRL)¹ and a regional library consortium called the Greater Western Library Alliance (GWLA)², it was agreed that a joint project would be formed to explore and pilot the development of the Technical Report Archive & Image Library (TRAIL).³ The project would be supported and managed under the 32 member institutions of the Greater Western Library Alliance. The charge for TRAIL was to:

*Develop a collaborative project with the Center for Research Libraries to identify, digitize, archive, and provide persistent and unrestricted access to federal technical reports issued prior to 1975.*⁴

A ten-member taskforce was appointed and an initial meeting was held in the fall of 2006. Ideas for workflows and costs of digitizing and hosting digital copies of technical reports were discussed. Throughout the months that followed discussions on the many pros and cons regarding collection acquisitions, workflows, digital conversion processes, digital archiving, and print archiving ensued.

From these discussions, the guiding principles for TRAIL were developed:

1. TRAIL would have a primary digital collection (and archive) with a corresponding print archive.
2. Opportunities for participation and involvement should be available to all institutions.
3. Efforts would be made to digitize and archive all items in a selected report series, before working on additional series.
4. The taskforce would only focus on reports issued in print format.
5. The taskforce would only focus on documents in the public domain or for which they received explicit copyright permissions from the copyright holder.

The taskforce started to work on digitizing its first collections. By the Spring of 2007, a working pilot collection was released. Upon review, GWLA agreed to fund the project for a year to continue the development of TRAIL. GWLA members each agreed to pay \$4,000 dollars providing TRAIL with \$128,000 to cover digitization, overhead and staff costs for the year. During this time, a number of partnership opportunities were presented that greatly aided the project in working towards its objective. TRAIL partnered with the University of Michigan (UM) and Google through the University of Michigan's Google Books project to aid with the scanning and digital hosting and archiving of many reports. Items that met the Google scanning requirements were cataloged, barcoded and scanned through this process. A year later when the University of Michigan and the Center for Institutional Cooperation (CIC)⁵ released the

Hathi Trust Digital Archive⁶, these items were deposited by the University of Michigan, as part of their collection. Items which cannot be scanned through the Google Books because of fragility, inclusion of maps or other special handling needs are outsourced to a digitization vendor. These items are then hosted and made available through the TRAIL public interface.

With the UM partnership came many changes to our processing workflows, one of which was creating a MARC record for each individual report processed. The benefit of this workflow change was that the project could then work with OCLC to provide collection sets for all of the digital versions of these items. Libraries could then purchase sets of records to be added to their OPACS increasing access to these report collections. (To date, many of the individual item records are searchable and available in OCLC. However, they have not yet been pulled into electronic collection sets by OCLC. This is due to the of needing to convert the OCLC records for the print items into new electronic records with the receipt of the URL from UM and Google. It is hoped that this will be resolved January 2009).

The work of TRAIL is overseen by the TRAIL taskforce. As a new report series is identified to be acquired and digitized a lead organizer is appointed. (It is not necessary for the lead to be a member of the taskforce.) This person is responsible for identifying and acquiring two complete sets of reports for the given series. (Items are contributed by libraries and organizations nationally.) One set will be used for digitization and the other set is secured by the print archive. As items are acquired they are sent to the University of Arizona Library (also referred to as the "Central processing site") where the items are checked for special format needs and copyright compliance. They are then catalogued (using either copy cataloging or original cataloging) and packed for shipping to either Google or another digitization vendor. UAL then oversees the digitization status and quality checks for all scanned items. As a report series are completed, UAL works with OCLC to close out the cataloging authorizations and create the collection sets.

Future Directions

After two and a half years (not including the pilot), TRAIL has digitized more than 15,000 technical reports, which is more than 1.2 million pages of federally funded research. TRAIL has received requests and comments from all over the world from researchers ranging from Chile to Greece whom otherwise have had difficulty locating, ordering and purchasing reports.

At present, the digitized collection is being moved from the pilot site to a more permanent location, which is split between the University of North Texas and the Hathi Trust (via partnership with the University of Michigan). A new public interface and search service is being created by the University of Washington. The new interface being developed will allow for a federated search of servers in both locations. This will greatly improve access to the more than 15,000 reports that are currently available.

The success of TRAIL continues to bring new opportunities. Most recently the TRAIL taskforce is exploring moving TRAIL under the umbrella of the Global Resources Network (GRN)⁷ overseen by the Center for Research Libraries. Moving to the GRN will allow greater ability to explore partnerships at a global as well as national level to partner on making technical research reports more widely available. It will also allow institutions to provide funding support and participate in the program without having to be a member of either the Greater Western Library Alliance or the Center for Research Libraries. Under the arm of the GRN, TRAIL will be better positioned to advocate on the importance of ensuring access to these important documents in the history of our research and technology development in the US. Individuals or institutions wanting to learn how to get involved with the Technical Report Archive & Image Library can find additional information at <http://trail.gwla.org>.

References

- ¹ CRL, Center for Research Library <http://www.crl.edu>
- ² GWLA, The Greater Western Library Alliance <http://www.gwla.org>
- ³ TRAIL, Technical Report Archive & Image Library <http://trail.gwla.org>
- ⁴ The Greater Western Library Alliance <http://www.gwla.org/technical-reports-archives-and-image-library-trail>
- ⁵ CIC, Center for Institutional Cooperation <http://www.cic.net>
- ⁶ Hathi Trust Digital Archive, <http://www.hathitrust.org>
- ⁷ GRN, Global Resources Network <http://www.crl.edu/grn>



TWELFTH INTERNATIONAL CONFERENCE ON GREY LITERATURE



web qr code

National Technical Library (hereafter referred to as “NTL”) is central professional library open to public, which offers unique collection of 250 thousand publications freely accessible in open circulation. Its holdings form the largest collection of Czech and foreign documents from technology and applied natural sciences as well as associated social sciences. It contains total of 1,2 Mil. volumes of books, journals and newspapers, theses, reports, standards, and trade literature in both printed and electronic forms. Besides its own collection, parts of Central Library of CTU in Prague and Central Library of ICT holdings — are accessible in NTL.

**Held in the National Technical Library in
Prague, Czech Republic, on December 6 – 7, 2010.**

As corresponds to its statutes the NTL runs – among others – the project of building the **National Repository of Grey Literature**. The project aims at gathering metadata and possibly full texts of grey documents in the field of education, science and research. The NTL supports an education in the field of grey literature through annual seminars in the Czech Republic.

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library

For the latest news on GL12 or National
Repository of Grey Literature, visit our
Web site at <http://nrgl.techlib.cz/>

Digitizing Grey Literature from the *Antarctic Bibliography* Collection *

Tina Gheen and Sue Olmsted
(United States)

Abstract

In 1962 the National Science Foundation (NSF) created a clearinghouse for Antarctic information intended to foster the global, free exchange of data and publications between scientists and researchers. With funding from NSF, the Library of Congress began assembling the Antarctic Bibliography in 1963, and full-text of the items listed in the bibliography was later captured on microfiche for preservation. The Antarctic Bibliography primarily consists of journal articles, monographs, technical reports and conference proceedings collected by the Library of Congress; however, other polar libraries contributed a generous amount of grey literature material to the bibliography as well. Only a few copies of the microfiche collection were created and distributed, so the full-text version is a rare resource preserved in an increasingly less accessible format. The goal of the Polar Digitization project at the National Science Foundation Library is to make the full-text grey literature materials from the Antarctic Bibliography microfiche collection, including unpublished works, materials from conferences that are not readily available and rare government reports, available to the public electronically on the web through digitization and cataloging. This paper addresses the challenges and importance of making the full text of grey matter from the Antarctic Bibliography accessible in digital form.

Introduction

The U.S. National Science Foundation (NSF) is a government grant agency which funds research and education in most fields of science and engineering. NSF's Office of Polar Programs (OPP) manages and funds the United States Antarctic Program (USAP)¹. In 1962, USAP established a clearinghouse for Antarctic Information at the NSF. The Antarctic Bibliography was a product of this initiative. From 1963 to 1998, the Library of Congress prepared the *Antarctic Bibliography*, a multi-volume set of citations and abstracts encompassing the world's scientific literature regarding Antarctica and the Southern Ocean². The bibliography coverage spans from 1951 to 1998, includes Antarctic biology, geology, engineering, medicine, meteorology, oceanography, atmospheric physics, geophysics, and political and social science. The bibliography contains two sets of information:

- 1) bibliographic records with very brief abstracts of the world's Antarctic research literature from 1962 to 1998 in all of the scientific disciplines relevant to Antarctica (approximately 65,000 titles) and
- 2) bibliographic records without abstracts of the world's Antarctic research literature from 1951 to 1961 in all of the scientific disciplines relevant to Antarctica (approximately 5,000 titles)². All material included in the printed bibliography was indexed by a team of subject specialists in the Science and Technology Division of the Library of Congress.

In addition to the printed volumes, microform copies of the full text of most of the cited items in the bibliography were produced by the Library of Congress as part of the funded project. These microfiche contain scientific articles, conference proceedings, technical reports, official memoranda, policy documents and other relevant publications. NSF has microfiche sets of the full texts of Antarctic items for on-site use only at the NSF building in Virginia, its three research stations in Antarctica, and the University of Canterbury in New Zealand (for NSF's Antarctic facility in Christchurch)³. The American Geophysical Institute (AGI) also has a set of the fiche at their headquarters in Alexandria, Virginia.

The National Science Foundation Library is an information and reference resource supporting the USAP and the NSF OPP. As custodians of one of the rare archival sets of microfiche of full-text materials collected from the *Antarctic Bibliography*, we have become increasingly aware of the need to make the grey literature information

* First published in the GL11 Conference Proceedings, March 2010

contained on this fairly inaccessible microfiche collection more readily available to the scientific community. The bibliography holds a wealth of historical information related to scientific data, climate studies, environmental changes, policy, and historical events. The technical reports alone cover topics such as snow road construction, seismic wave dispersion, studies of ice cores, geodetic surveys, chemical profiles of ice shelves, marine biological studies, and cartography. Many of these technical reports and government-funded studies are not accessible and are no longer available as monographs. In addition, the fiche version of the bibliography captured proceedings from societal conferences held all over the world.

The *Antarctic Bibliography* is currently maintained and updated online by the American Geological Institute (AGI)⁴. AGI distributes the bibliography free on the web through the Cold Regions Bibliography Project (<http://www.coldregions.org/>) and provides links to full-text when possible, but much of the full text grey literature must be digitized on demand for a fee. The National Science Foundation Library began a pilot project to digitize the grey literature contained on the *Antarctic Bibliography* microfiche. However, digitizing and delivering the content currently preserved on microfiche presents a host of challenges including material selection, language barriers, original image quality, inclusion of metadata, and copyright issues.

Methodology

Much of the material contained in the *Antarctic Bibliography* was acquired from journals and books. Grey literature can be found throughout the collection, particularly in the materials submitted by polar libraries. Because much of the information captured during the microfilming process is now copyrighted and available from journal publishers, the selection of materials to digitize becomes critical. Therefore, we concentrated on the evaluation of technical reports, rare conference proceedings, and other information which falls into the category of grey literature. For this paper, we define grey literature as information that is not widely published or distributed, and which is usually difficult to discover or obtain, particularly unpublished works, materials from conferences that are not readily available and rare government reports.

Because NSF does not have the necessary equipment or staff to produce quality digital reproductions of the fiche, samples of the fiche were sent to the OCLC Preservation Service Center for automated digitization. The deliverables for the project included uncompressed 300 dpi grayscale archival TIF files and derivative bitonal bound PDF files with hidden text.

For our pilot study, we selected approximately five hundred fiche (~3800 individual cells) of grey literature materials from the full-text version of the *Antarctic Bibliography*. The samples were selected based upon a manual, internal evaluation to determine whether the fiche contained grey literature materials and whether the materials existed in digitized form on the web.

Anatomy of the Antarctic Bibliography Fiche

Manual identification and selection was a major, time consuming component of the initial phase of our pilot. The full text materials of the *Antarctic Bibliography* are not often easy to distinguish separately. An individual fiche may contain only one article, book, or proceeding, or it may contain a combination of several articles, books, and reports. Careful manual inspection for a black cell or unique identifier is usually required to distinguish one item from another. Similarly, a packet of several fiche may contain one long article, book, proceeding, or report, or a packet of fiche might span a number of individual publications.

Since our project specifically targeted grey literature, we did not provide a single block of hundreds of consecutive fiche because of the likelihood the block would contain mass produced full text books, commercial online journal products, and other materials which involved copyright issues with publishers. The consecutive packets often contained some desired technical reports and proceedings, but also included one or more journal articles, which were already digitally available from a publisher. Because of this, it was often impossible to select fiche that contained only grey literature publications. It was likely materials which had already been converted to digital format would be included in the span.

Evaluation and Selection

In order to avoid duplication of efforts and best use the financial resources available to us, we needed to carefully evaluate and select sample items to send to the preservation specialists at OCLC. Using the index of the *Antarctic Bibliography*, staff selected desirable samples of technical reports, proceedings, and other materials. These samples were identified by a unique accession number from the Library of Congress, which was added to the *Antarctic Bibliography*. Staff then manually recorded fiche number, item type, author, article title or name of conference, number of pages, and publication year. This information was used in searches on various online databases such as OCLC WorldCat and Public STINET to determine if the material was readily available. If the selection was available from several lending institutions or agencies, it was not a candidate for our pilot.

After evaluating the accessibility of the samples, staff manually assessed and evaluated the fiche. The number of cells per fiche, the number of images per cell, title, and number of owners on WorldCat were recorded. Staff then compiled a master list which contained:

- Library of Congress accession number (ID #)
- Corresponding volume of the print version of the *Antarctic Bibliography* where the citation was found
- Material type:
 - Book
 - Book Chapter
 - Booklet
 - Information Bulletin
 - Proceeding
 - Report
 - Technical Note
 - Technical Report
 - Etc.
- Number of images contained in the sample
- Publication year
- Additional Comments
 - Language
 - Availability

Digitization

We analyzed this data to choose a variety of appropriate grey literature materials for our pilot in one small sample set, which was sent to the OCLC Preservation Service Center in Bethlehem, Pennsylvania. From these samples, OCLC created a single TIFF per fiche frame. Because the process is automated, target or divider pages could not be excluded and were also digitized. This smaller initial sample set allowed us to evaluate quality before continuing with more digitization.

Analyzing and selecting specific publications led us to discover issues related to the initial filming of the microfiche. Nonuniformity in microfilming and standards created challenges for automated digitization processing. Fiche were filmed by various technicians using different standards, equipment, and procedures over the years. In particular, the background used for some of the fiche appeared to be a highly textured fabric. This background caused problems for the automated OCLC equipment since automatic detection of the page edge was impossible. To solve the problem, the fiche were scanned using an automatic crop box. Unfortunately, some of the pages were so badly skewed during the initial fiche filming that part of the page fell outside the cropped area (Image 1).

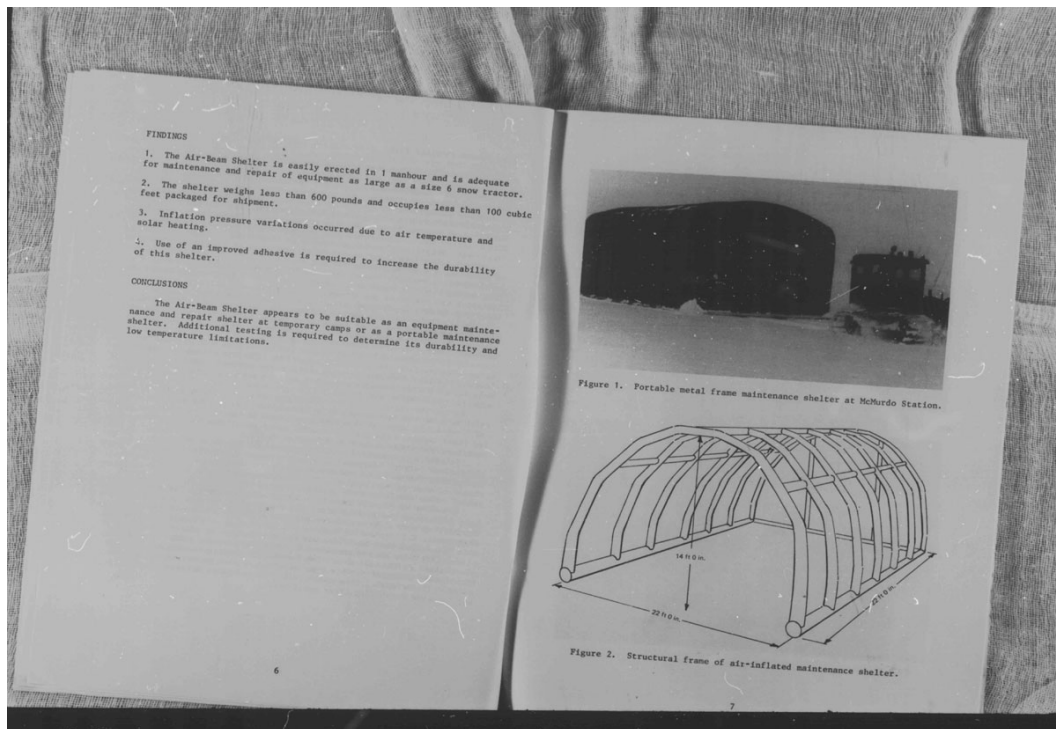


Image 1. A scan from the *Antarctic Bibliography* microfiche showing the textured background and skewed filming⁵.

Rigid human quality control at the OCLC Preservation Service Center was required to remedy the problems with skewing and background noise. The resulting images were far better (Image 2).

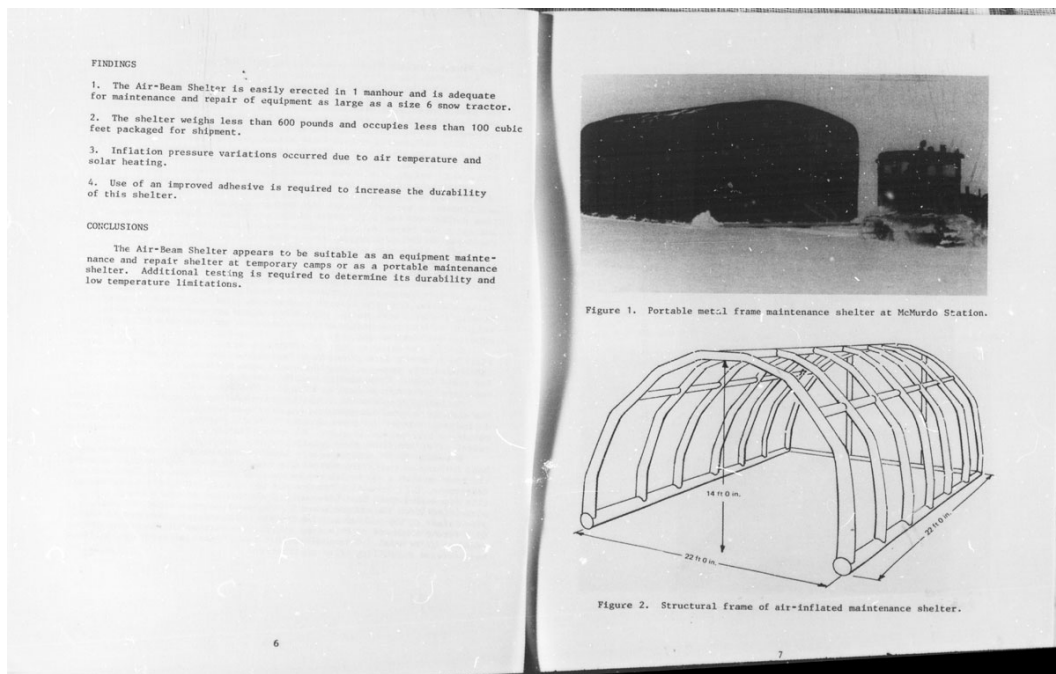


Image 2. Final scan from the *Antarctic Bibliography* microfiche⁵.

Additionally, there was no standard for the number of pages included per frame on the fiche. Some fiche contained two images or pages per frame. Other fiche only contained one image per frame. Individual frame widths varied by several millimeters. Additionally, some fiche contained thirty-six frames per fiche, while others contained sixty to one hundred frames per fiche.

Ideally, we wanted a digitized image for each individual page, rather than a digitized image for each frame. However, after working with OCLC to find a solution, we realized we could not achieve one page per image with their automated process because of the issues and inconsistencies noted above. The OCLC automated process could produce a single TIFF per fiche frame up to 17" X 11". Because our goal was to create a quality digitized image that could be scanned using OCR technology for accessibility, we decided to forsake one image per page for now. We again recognized additional library staff time would be required to separate the pages of the digitized documents included in a single frame.

Results

OCLC delivered a hard drive containing a little over 2500 TIFF files containing one frame per TIFF. This translates to around 4500 pages and approximately 77 articles. The quality of our fiche images was poor, so OCLC advised against performing OCR and converting to PDF.

Conclusions

We learned much from our pilot. Sufficient time and a defined protocol for the selection of materials for digitization are critical. Although we spent a great deal of time evaluating the fiche and considering the best methodology, more time dedicated to this part of the process would have been beneficial. It is important to carefully assess the quality of the material to be digitized. Automated digitization systems have limitations and materials should be evaluated to determine if an automated process is a viable option. In our case, manual scanning would perhaps be more cost effective in the long run. Human quality control incurred higher costs over the course of the pilot, but proved to be crucial to achieving usable results. Because of this quality control process, digitization was halted so we could address issues with image quality and fiche composition.

Communication with peers in the field is vital to avoid duplication of effort. We were conscious of the need to avoid duplication of effort. The information regarding digitization projects and availability of materials we received from AGI and other libraries was useful and helped shape the direction and focus of our pilot.

Future Work

Because the quality of the images is poor, we believe image preprocessing is required. OCR technology works best on clearly scanned pages without borders, since most of the content is composed of easily recognized characters. However, the fiche images often contain background content outside of the page itself, which OCR does not deal with gracefully. A multi-pass filter approach, which detects edges, normalizes the image, and removes noise, is an optimal solution. The resulting data analyzed by the OCR program would be much cleaner and easier to process. OCR performs the first step - digitization of the grey literature contained in the fiche. The next step is to transform that data into something searchable and machine readable. Conversion of the metadata into XML would be a good first step towards this goal. Unfortunately the fiche does not contain a good amount of metadata, so metadata creation must be a manual process performed by subject matter experts. The XML can then be used to create a taxonomy to classify the articles and data contained in them. To further augment this, we could create an ontology based upon the taxonomy to make the materials semantically available and searchable.

In the future, we hope to secure additional funding or partner with another organization to continue digitizing the rare and valuable full-text found in the *Antarctic Bibliography* microfiche. Once this grey literature has been digitized, processed, tagged with metadata, and cataloged, the NSF Library intends to collaborate with AGI to make the materials available free of charge to the public on the web. Ultimately, this project will result in increased access to the grey literature recorded in the *Antarctic Bibliography*, which in turn we hope will contribute to greater scientific discoveries, historical revelations, and better education about Antarctica and the Southern Oceans.

Acknowledgements

We respectfully thank Guy Guthridge of the National Science Foundation and Sharon Tahirkheli of the American Geological Institute for sharing their knowledge and experiences with us.

References

1. National Science Foundation (US). Investing in America's Future: Strategic Plan 2006-2011, Arlington, VA: National Science Foundation (US). (2006).
2. Office of Polar Programs. *Operation of the Antarctic Bibliography and the Bibliography on Cold Regions Science and Technology. Program Solicitation NSF 99-94*. Arlington, VA: National Science Foundation (US). (1994).
3. National Science Foundation (US). *Questions and Answers, Program Solicitation NSF 99-94*. Arlington, VA: National Science Foundation (US). 1999 Apr [Cited October 21, 2009] Available from: <http://www.nsf.gov/pubs/1999/qa9994/qa9994.doc>
4. Office of Polar Programs. Frequently Asked Questions about the U.S. Antarctic Program. Arlington, VA: National Science Foundation (US); 2008 Jul [Cited October 21, 2009]. Available from: <http://www.nsf.gov/od/opp/frequent.jsp>
5. Brier, F.W. *Air inflated portable maintenance shelter*: Technical note, N-1395. Port Hueneme, CA; U.S. Naval Construction Battalion Center. (1975).

Web 2.0 Utilization in E-Diplomacy and the Proliferation of Government Grey Literature

Jonathan Omowalé Cain
(United States)

Introduction

This paper explores the use of Web 2.0 technologies in the creation of Grey Literature as it functions in the United States Government. The role of Grey Literature has developed significantly with the growth of e-government and particularly during the Bush Administration. The foreign policy of the Bush Whitehouse, expressed through Secretary Condoleezza Rice, was one of Transformational Diplomacy. In a speech at Georgetown University in 2006, Secretary Rice explained "Perhaps the newest and most cost-effective way to adopt a more local posture is through a Virtual Presence Post. Here one or more of our young officers creates and manages an Internet site that is focused on key population centers. This digital meeting room enables foreign citizens, young people most of all, to engage online with American diplomats who could be hundreds of miles away."¹ In line with this push to engage in e-diplomacy, the growth of the Internet has provided the opportunity for academics, researchers, and activists to voice opinions instantaneously. The transition of power between White House Administrations (from Bush to Obama) brought a greater emphasis on the utilization of third party applications (YouTube) as opposed to closed network technological solutions (Diplopedia²). These third party resources oftentimes do not have the same obligation, nor desire to preserve or index this material with the consistency that the federal government, or scholar/historians would like. On January 3, 2005, Mary Mosquera wrote an article entitled "GPO Seeks Tools to Harvest Overlooked Documents on the Web" for *Government Computer News*, that hinted at this problem of information regulation in a digital world. The article pointed out that "Federal agencies are increasingly publishing information only in electronic formats and frequently fail to inform GPO of new publications that should be included in the depository library and cataloging programs" and explains that the "GPO wants to use Web crawler and data-mining technologies to retrieve publications from Web sites to identify those that agencies have not catalogued for its Federal Depository Library Program and the Cataloging and Indexing Program"³.

This paper will investigate the functionality of higher government models regarding e government and the ways in which they produce information for intragovernmental communication as well as external consumers of information.

Methodology

eDiplomacy is a new and emerging field of study and utilization. When approaching this paper, I had formulated an idea about what eDiplomacy was, but had no concrete leads. In order to gain a more adequate understanding of how eDiplomacy is used, I began by exploring the concept of eGovernment. eGovernment was not the most fruitful avenue of search when trying to understand the ways in which the government uses the internet to communicate. eGovernment is first and foremost concerned with methods of communication between government and its citizens/subjects for information distribution and subject participation. Studies of eGovernment lead me to explore the UN document on eGovernment readiness, entitled *UN E-Government Survey 2008: From E-Government to Connected Governance*. The UN's focus governing through wired technologies reoriented my concentration to the recent change in Presidential administrations. During and immediately following the 2009 elections, the U.S. media discussed the effect of new technologies on the campaign trail. Researching the correlation between communicative technologies and the Obama administration led me to the *Putting Citizens First: Transforming Online Government*, A White Paper Written for the 2008-2009 Presidential Transition Team by the Federal Web Managers Council November 2008. Discovering this paper, gave way to the idea to search for more documents published by the Federal Web Managers Council. A select group of the council published *Social Media and the Federal Government: Perceived and Real Barriers and Potential Solutions*. This document informed much of what I came to research in understanding the government's publication through nontraditional publication arms (such as blogs). The first point of discussion was "The context for using social media within the federal government"⁴. This took the emphasis of my study from an examination of opaque communication

tools (like secure wikis), to one that includes a thorough examination of the diverse publication venues that the diplomatic organ of the United States uses.

In studying the way that government is embracing web 2.0 technologies for different tasks of diplomatic work, I first explored the ways that wikis are being used for research and knowledge sharing in the internal organs of information gathering with both *Intellipedia* and *Diplopedia*. Intelligence/information gathering is a critical function of diplomatic work so I incorporated the tool of the intelligence community (*Intellipedia*) along with the State Department's tool (*Diplopedia*). Another aspect of diplomacy pertains to communicating the message of your nation abroad. To gauge the ways that these new services affected the publication of State Department materials, I explored the official website (*State.gov*), the official blog (*Dipnote*), as well as the twitter, flickr, facebook, and YouTube accounts. I examined each service, gauged the type of material made available, the frequency, as well as the popularity and potential impact of each service to produce government information.

Understanding eDiplomacy

In October 2003, The Office of eDiplomacy was created as a part of the Bureau of Information Resource Management. The office of eDiplomacy was created in an effort to improve communication and knowledge sharing in the diplomatic core. According to its official about page at www.state.gov, there were "three broad mandates" that guided the creation of eDiplomacy in the United States Government: "Bring State diplomats into the Information Technology decision-making process; improve ways to connect to and collaborate with others inside and outside State; and promote knowledge management."⁵

The concentration of internal communication and sharing is reflected in the major programs in which the Office of eDiplomacy is engaged including, *Diplopedia Wiki*. The *Diplopedia* wiki is "the internal unclassified online encyclopedia" used "to create a broad, informative and expanding reference tool or knowledge-sharing about the Department, its programs and offices and other international affairs subjects"⁶ (state.gov/m/irm/ediplomacy/c23480.htm). The second major project that was launched under this initial drive was "Knowledge Leadership." The term is a derivative of the library, enterprise discipline of Knowledge Management. Again this is an inward facing and affecting project. It was created recognizing the weakness in the lines of information from disparate information gathering outposts especially with the growing concerns surrounding terrorism. The goals of the focus on four areas: "Use of self-forming, self-managing online communities for collaboration across geographic and organizational boundaries; Better access to knowledge resources; Better access to expertise; Widespread use of IT tools that build knowledge-sharing into the daily work flow"⁷.

As opposed to eGovernment initiatives that work to improve the amount of information and participation among everyday citizens, these initial approaches to eDiplomacy are concentrated upon inter departmental communication. This understanding of what it means to be engaged in eDiplomacy would soon change, as administrations changed. When Barack Hussein Obama was elected the 44th President of the United States, it became a pronounced goal of his administration to bring new web 2.0 technologies to government transparency. This emphasis on web based, communicative "new technologies" spread even to the diplomatic organs of the United States, as a result you see a growth of multimedia and social networking technologies.

While exploring the concept of diplomacy in the digital realm it is easy to focus on the initiatives of the State Department, but it is important to recognize the sea change that the development of the digital world has brought to communication in general and to the world of diplomacy in particular. When exploring the impact of the Internet upon information and communication it is easy to become idealistic. But it is key to remember that information is accessible to a far wider community, Nicholas Westcott points out it "has a dark side, illustrated by its use for fraud, pornography, and terrorism. The Internet is a vehicle in which eBay and Al Qaeda are fellow travelers."⁸ Despite the dark aspects of the internet, it has been a very instrumental tool in engaging a wider community beyond the state in the process of international diplomacy, "NGOs have been swift to adapt the potential of the Internet to increase their influence in international affairs."⁹

Department of State

The Department of State was the first federal agency created under the United States Constitution. Its early creation is an expression of its importance to the United States government. It is the outward face of the United States to the world. In that capacity, capturing and exploiting the new communications opportunities presented by the advent of the widespread broadband was a necessity of the State Department.

General Web Presence:

With www.state.gov, the State Department has wholeheartedly engaged in the public information aspects of eDiplomacy. On the website you will find the traditional features of a promotional website. With minimal effort, it is easy for users to discover: who the Department of State is, what the mission statement is, where it operates and how it does so. The information that is produced and housed on the site and its auxiliary digital repository (archive) can easily be understood to be grey literature. The Grey Literature Network Service defines Grey Literature as "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e., where publishing is not the primary activity of the producing body".¹⁰ New, original and timely information is consistently uploaded, published, and housed on the site; documents that may be difficult, or in some cases impossible to locate through other channels. This approach to publication is not unique to the United States government, more interesting the United States Department of State uses State.gov in a similar manner as some universities are using institutional repositories, as informal or nontraditional venues of publication.

The U.S. Department of State produces a large quantity of grey literature primarily through its website, and continues to make it available through the associated archive. The page itself produces a great amount of mixed media publications, including text and video components. Navigating the text available on the live page is simple. There is a plain text search box, for simple searching and a link to a subject index. Underneath the main navigation bar you will see a "find it by:" bar. Here you are given the option of locating material by "topic," "country," "publication type," "speaker," and "type." There are some obvious limitations to the date search, the options available include: "most recent," "past 7 days," and "past 90 days." Therefore older documents do not appear to be available through this method of search. In addition, a cursory search of the "publication type" shows you what type of material is indexed on the site. These materials include: articles, background notes, communiques, daily appointment schedules, daily press briefings, fact sheets, funding opportunity announcements, interviews, op-ed pieces, press releases, remarks reports, and testimonies. Materials such as videos and photographs are accounted for among the resources in addition to the text pieces. To explore the wider variety of the materials that are produced/published by the Department of State one must search the "multimedia," a different avenue of information created by the State Department. It is interesting that these are not considered as publications by the Department of State.

Wikis:

Wikis were among the first initiatives that addressed the goals of Transformational Diplomacy that were established by the Bush Administration. In this paper two main wikis will be examined, Diplopedia and Intellipedia. Both of these wikis were designed for the transfer of information within the government, and in rare cases, with external partners as well. The wiki belonging to the Department of State, described previously, was announced at Wikimania in 2006. Career Diplomat Robert Bronk, presented Diplopedia to the world and its basic functionality in a talk entitled *Diplopedia: Application of the Wiki Model for Collaborative Drafting in Foreign Affairs*.¹¹ The collaborative information system is being viewed favorably by foreign diplomatic organs as well. In *Citizens at the Centre: Public Engagement for Better Policy and Services*, it is reported that the "use of online collaborative tools has helped to foster communities of interest among State Department employees posted all over the globe"¹². This concentration on information sharing fulfills the information sharing and content management goals of the Federal government. In the updated Department of State's Information Technology Strategic Plan, an emphasis is placed on "Collaboration tools that support self-forming, self-managing professional networks and communities of practice and interest that overcome boundaries of geography time and organization".¹³ In addition, to this internal information reorganization, there is also effort to share this information with partners. In its strategic goals for 2006-2010, the Office of eDiplomacy states "The Department works in concert with other U.S.

Government Agencies, Non-Government Organizations (NGOs), and foreign governments in promoting the U.S. foreign policy agenda."¹⁴ This cooperation is a reflection of the ways in which the growth of the Internet has rewritten the rules of diplomacy and the cost of participation. In *Digital Diplomacy the Impact of the Internet on International Relations* Wescott posits "... studies of the Internet and diplomacy have identified that the Internet enables more and different actors to get involved in political and diplomatic processes." He points that in recent history, outcomes of international political events, have been "...broadcast not only through Government channels but across the Internet by multiple NGO websites."¹⁵ Intellipedia, a product of the Central Intelligence Agency (CIA), has been described as "a version of the well known Wikipedia for spies".¹⁶ The May/June issue of *Pathfinder: The Geospatial Intelligence Magazine* discusses the importance of Intellipedia and similar products internally and in the intelligence community at large. In *Collaborate, Cooperate, Integrate* a column penned by Mike McConnell (Director of National Intelligence) some interesting figures shed light on the importance of the resource to the intelligence community at large. "The number of NSA self identified Intellipedia users has jumped more than 81 percent in the last six months with over 400 distinct article pages posted."¹⁷ The value of these open styled closed resources is being recognized in various government communities; whether it is through Diplopedia (Department of State) or Intellipedia (Intelligence Community) government actors are using these grey literature channels to "allow groups of people who otherwise would have only limited contact with each other to pool their knowledge in a single repository available to all within that community."¹⁸ Beyond exploring the perceived impact of the Diplopedia there is little to discuss. Information on these wiki platforms are not widely accessible to the general public.

Evaluation of Social Media and eDiplomacy:

The Federal Web Managers Council, in the article "Social Media and Federal Government," assert that "we've seen that social media in government has become the number one topic of discussion within our government web manager community."¹⁹ For the government at large, the introduction, and proliferation of these resources in everyday life raise a myriad of questions for information gathering and diplomatic endeavors. "The new administration should require agencies to provide access to social media sites unless the agency head justifies blocking certain employees or certain sites."²⁰

Despite the fact that the majority of the materials discussing the utilization of web 2.0 applications as well as open source software is concentrated upon notions of consumption of these new technologies, there is some regard given to exploitation of the resources. For instance, in *Social Media and Federal Government*, the corporate authors suggest "the government shouldn't arbitrarily decide which companies will be given the cachet of providing our content, which can provide value to the value to their sites. For example, federal agencies should have criteria to determine which video sharing sites they will publish their videos to." They continue "Employees with a clear business need can create accounts to use free services, as long as they have managerial approval."²¹ It is in this point that we can see a forward thinking movement regarding tactical utilization of web 2.0 resources. The State Department with the change in administration was among the first to use these new tools.

Blogs:

Julian Mahler and Priscilla Regan conducted a study on the government's use of blogs in *Issues in Governance Studies*. They explored the ways that major government departments used blogs to promote a message, and potentially produce a measure of community. The issues covered, such as the mediation of information between individual and federal agency are important considering the manner in which web 2.0 technologies fulfill the niche of grey literature production. For the purposes of this paper the question is not heavily engrossed in the functions of eGovernment, i.e., individual participation in policy making, but rather the dissemination of legitimate, if not traditionally published, information for diplomatic purposes.

Dipnote²²:

Dipnote is the official blog of the United States Department of State. The Dipnote editors explain "Blogs.state.gov offers the public an alternative source to mainstream media for U.S. foreign policy information. This blog offers the opportunity for participants to discuss important foreign policy issues with senior Department officials."²³

There is a wide variety of information available on Dipnote. Dipnote serves as a publishing forum and a place to outreach and defend foreign policy. In addition to the

traditional posts by the authors of Dipnote, videos and photographs are featured as well. The breadth of material covered on the blog includes an actual representation of timely issues of United States policy (What Issue Is Most Critical to a 21st Century Partnership Between the U.S. and India?²⁴), to recent news (Dr. Shah's Statement²⁵), and pieces dealing with the quotidian aspects of foreign service life (Thanksgiving Without Borders). This blog also seems to generate a dialogue of user commentary; these comments are not an interdepartmental collaboration, nor do they seem to be a representation of governmental dialogue with nongovernmental actors. Nevertheless, the blog is representative of a federal agency producing content on a regular basis. According to *Blogs as Public Forums for Agency Policymaking*, in March 2008 the authors counted 16 postings, and 167 comments.²⁶ The lack of comment activity is relatively surprising. Thinking that the reason for the low level of communication among the commenter and staff may have something to do with the rules governing commenting policy, it is relatively surprising that the level of scrutiny is extremely lenient. In addition to the featured content, the blog also links to other government blogs and affiliated (nonofficial State) blogs, as well as RSS features and twitter updates.

flickr²⁷

Another official publishing venue of the Department of State is flickr(TM). flickr is representative of the way in which official government agencies are using web 2.0 technologies to present pertinent information to a wider community. The collection on flickr "represents a special collection gathered by the Department's Office of the Historian. The Office of the Historian is responsible, under law, for the preparation and publication of the official historical documentary record of U.S. foreign policy in the Foreign Relations of the United States series."²⁸ These non traditionally published collections are open to the users like all other photostreams on flickr. The State Department offers users the ability to interact with the collections through adding tags and comments. The comments feature appears to be sparsely used but offers users, American citizens or foreign citizens to view the published activities of the Department. According to the flickr map the Department of State's account features approximately 1,081 items.²⁹ Despite the efforts, it appears that the State Department has yet to fully populate their flickr space. Thus far, it has yet to establish either "galleries" or people tagging. Although the account has been in existence and populated only since 2008, the collection has information stretching back periodically to 1902.

Facebook³⁰

The U.S Department of State maintains several pages on facebook. The primary page on facebook provides many of the same stories that one would find on the official website or even the Dipnote blog. Despite the similarity of the availability of information there is disclaimer "Welcome to our Facebook page. If you're looking for the official source of information about the U.S. Department of State, please visit <http://www.state.gov>"³¹ This is a place where you will notice there is a greater level of interaction among the community here. On flickr and even on Dipnote, most of the posts or photos would only have periodic comments, the posts on facebook seem to be better received and more frequently commented upon. It may be the nature of the technology, or even the extreme proliferation of facebook in everyday life. The discussion section, provides a space and an impetus for communications between nondepartment actors and state employees to dialogue in a more natural fashion than in Dipnote, which is more of a publishing venue, or even flickr which serves in the same capacity.

As with the nature of facebook, the State Department offers a variety of pages reflecting the varied facets of the Agency include the following.

COP 15 UN Climate Change Conference 2009³²

According to the synopsis, "This page will provide news, updates and information about U.S. activities leading up to and during COP-15."³³ Furthermore, the authors of the page implore viewers to "Check back often!"³⁴ This statement itself expresses the position of the Department of State view on facebook pages in diplomacy. The authors of the page, at the very least feel that this is an effective way to communicate the views and activities of the State Department at the Copenhagen Climate Change Conference. Through the posts, the government is seeking to publish its message on the transactions of the conference and the American interaction. What is unique to facebook, is the publication of third party materials. We have seen through Dipnote and twitter the State Department has chosen these publication venues to distribute

documents from the White House publications, but it is here that uniquely, State is publishing articles from sources such as *Yahoo! News*, *The Washington Post*, and other news sources.

Careers: Careers in Foreign Affairs³⁵

This is an unique example of grey publications having a distinctive impact on international diplomacy. Through these pages, the State Department is able to make some inroads in the recruitment of future diplomats. The stated purpose of the page is to "... network with U.S. diplomats, recruiters, employees and current and former interns. You can learn more about the opportunities and the experiences of those who have chosen a career in the Foreign Service, Civil Service or as an intern."³⁶ Effectively, this page serves as an expansive, dynamic recruitment poster.

The site includes discussion sections where people, potential applicants, can discuss issues pertaining to service and engage the Department of State in questions and concerns about the profession and the process of application. In addition to these discussions, there are videos bearing the departmental seal, that offer information on the careers in the Department of State, as diplomats as well as in civil service life and career.

Civilian Response Corps³⁷

The civilian response core is an interesting page, not for the quantity, quality, or uniqueness of the material on offer. There is very little information offered other than the definition of the corps: "Civilian federal employees and volunteers from the private sector and state and local governments, who train to deploy rapidly to countries in crisis or emerging from conflict, in order to provide reconstruction and stabilization assistance..." and the disclaimer "NOTE: Comments by individuals of this group do not represent the official position of the U.S. Department of State."³⁸

Consular Affairs (CA)

The purpose of the department determines the content and features of the page. In the information box, the department claims "We deal with events and issues that have a personal impact: birth, death, marriage, adoption, child custody, citizenship, and relocation to another country. CA provides the passports that enable Americans to travel internationally and stands ready to lend a helping hand when citizens fall victim to crime, accident or illness in other countries, or just want to vote absentee. We make decisions and take actions every day that form key turning points in people's lives."³⁹ As a result, the page is not as content rich as the main facebook page for the Department of State. This page is not a truly effective page, and has very few publications. The most interesting thing about the page, is its substantial following (697 people), that would make it a potential place to distribute information.

Diplomatic Security⁴⁰

The page for the Diplomatic Security Services is a page with very little official information, considering its function as an insular behind the scenes function of the State Department "Diplomatic Security does not formulate foreign policy. It plays another essential — yet behind the scenes — role: To provide a safe and secure environment for the conduct of U.S. foreign policy."⁴¹ This makes its position on publishing reasonable. As a result the information made available on the page is of little consequence or value in the mission or official message of the Department of State.

Global Women Issues⁴²

The page describes itself along the lines of the of its department, with an emphasis on "Promoting women's social, political, and economic equality around the world."⁴³ This page serves in the capacity of a web publisher of the programs content. Material that is established pertaining to the programs goals, news and information that is deemed pertinent to the mission statement is published on this page with links to the originating document. On the main page, the posts stretch back with some regularity, to June of 2009. This suggests that the department is finding this as a viable way of distributing the information that it sees as pertinent to its message. For example, the page reposts information from varying news sources that are accessible to the wide publics such as the Associated Post, Yahoo News etc., and then comments on the content of the article report. The commentary is providing a type of policy statement on the issue, providing first hand interaction (on a policy level) with the site's participants. Furthermore, this is another place in which one can see the democratization of diplomatic work, in addition to the commentary provided by the

staff of the page readers are able to enter the dialogue of the issue immediately and directly, having the information they present displayed through the same channel as the Office of Global Women's Issues. In addition, the page provides an official biography of the Office head (Ambassador-at-Large Melanne Verveer), and an opportunity to dialog with her (or more likely her staff) directly via the dialog section of the page; this provides a difference in the way that diplomatic information and access has been made available in a pre- Web 2.0 environment. As of December 2009, the facebook page had 775 followers, many of whom comment, and discuss the information made available on the page.

In addition to the pages, there are links to 41 U.S Embassy Facebook Pages. Each of these pages present news, updates and assorted information about the efforts of the United States in each of these different missions.

facebook has proven to be an interesting method of communication and information channel for the diplomatic arm of the United States Government. In a way it provides the same features of a less engaging medium like a newspaper or a commercial, but has enriched the way information is distributed through this unique interface. The department can choose to publish whatever information it deems appropriate through these channels creating nontraditional and dynamic government documents. Interestingly enough, it also allows members of the facebook community, one that spans the entire wired world, to follow the information of the respective pages. In doing so, the State Department can push their message directly to the homepages and inboxes of a facebook facilitated community. In many ways, this service adheres to a definition of grey literature that demands it be outside the world of traditional publishers, oftentimes original, and typically current. The nature of information exchange in a web 2.0 world makes it all more likely that this is the case.

Twitter⁴⁴

twitter is a very interesting example of government publications. twitter is designed to provide brief updates and links to longer bits of information. The structure is predetermined by the format, the nature of the information is structured to 140 characters. Originally, twitter was designed to be a social networking services, a service of friends and associates communicating, a concept that has interesting ramifications for eDiplomacy perhaps, but not necessarily in the context of grey literature study. The original purpose of twitter was reinforced by its origin question, "What are you doing?" The administration of twitter noticed an interesting change occurring based on a early omission on their behalf. On the twitter blog, they write "To stay simple, Twitter did not require individuals to confirm relationships. Instead, we left things open...People, organizations, and businesses quickly began leveraging the open nature of the network to share anything they wanted, completely ignoring the original question, seemingly on a quest to both ask and answer a different, more immediate question, 'What's happening?'"⁴⁵. This change in philosophy transformed the service from social networking in focus to more of an information sharing network. Content providers, corporate and government alike took note, among them the United States Department of State.

The State Department on twitter serves to promote the updates from Dipnote, and push information pertinent to State's mission from other twitter accounts including the White House. Interestingly, on the main website the link to the State Department's twitter account leads you to the twitter Dipnote but, there is another account in the form of TravelGov. TravelGov⁴⁶ is the "Official Twitter for U.S. Department of State Bureau of Consular Affairs - protecting the lives and interests of American citizens overseas"⁴⁷. This pushes the most pertinent information for Americans abroad through twitter. twitter has proven to be a consistent and useful way to publish information, the message of the Department of State, and the administrations policy. The twitter associated with the State Department's page is currently followed by 9,875 people and has published over 2,371 tweets. The importance of twitter should not be undervalued especially when you evaluate the followers list. The followers include US embassies abroad; politically active individuals (ChampionDarfur); professional interlocutors (mc_ellison); professional media outlets (LATpolitics, HuffImpact); academic programs (JohnsHopkinsSPH, yalenurse); NGOs (GlobalHealthorg, sos4children) and foreign organizations (TheCIC, UNJusticeOrg) as well as foreign based private corporations (Radnon,TStweets). For these subscribers, the Department of State's twitter is an effective and timely way of receiving government information, and potentially influence the Department of State.

YouTube⁴⁸

YouTube is a video publishing tool for the U.S. Department of State. State.gov, the official website of the United States, publishes video created by and for the United States Department of State. The U.S. Department of State video channel (statevideo) contains "official video produced by the U.S. Department of State. Videos include remarks by Secretary Clinton, Daily Press Briefing, special video collections based on foreign policy issues, and candid interviews of U.S. diplomats."⁴⁹ After President Obama's recent policy speech on Afghanistan (December 1, 2009), Secretary of State Clinton produced a video addressing the Afghan people, *Secretary Clinton's Message to Afghanistan*⁵⁰, expressing the U.S. goals moving forward. This is an expressly targeted effort towards a foreign nation through diplomatic channels. This particular video is produced not only in English but in Urdu and Pashtu, therefore indicating that the State Department views this channel as an effective way of communicating outward to other nations.

The channel was created February 25, 2007, has over 3,000 subscribers and has been viewed 168,131 times. Thus far, 344 videos have been published through the channel covering a variety of topics grouped in the form of playlists. These playlists include: *Careers at the U.S. Department of State*, *Secretary Clinton*, *Videos from the Americas*, *Passports*, *Visas & Travel*, *Middle East Digest*, *Daily Press Briefings*, *In Other News* and *U.S. Department of State: Feature Videos*. On these sites for official statements, text of the videos can be reached through a link to state.gov.

Conclusion

As far back as the establishment of the constitution of the United States of America, the government has been dealing with the use of government documents. Article I section 5 clause 3 of the Constitution states "Each House shall keep a journal of its proceedings, and from time to time publish the same"⁵¹. This concern with publishing and distributing government information extended to the Government Printing Act of 1895 which "established the position of Superintendent of Documents in the Government Printing Office and provided that, he should have general supervision of the distribution of all public documents except those printed for official use of the executive departments and the two Houses of Congress."⁵² This preoccupation with making the information available continued apace with the expansion of government and the progression of technology. In the contemporary world, the development of inexpensive broadband Internet access has fostered the proliferation of varied communication mediums, the nature of publication has changed. The modern Department of State has embraced these technologies for internal and external communication. Condoleezza Rice, the former Secretary of State, oversaw the transition to a policy of "Transformational Diplomacy." The policy establishes a pronounced virtual presence in diplomatic missions abroad. The efforts on facebook, twitter, and other third party Web 2.0 technologies has proven to be a consistent and seemingly effective method of conducting this new diplomatic philosophy.

References

- ¹ "Rice | Remarks at Georgetown School of Foreign Service." http://www.unc.edu/depts/diplomat/item/2006/0103/rice/rice_georgetown.html (accessed 12/5/2009, 2009).
- ² According to, <http://www.state.gov/m/irm/ediplomacy/115847.htm> ; "Diplopedia is an online encyclopedia of foreign affairs information. It is a wiki, a kind of Web site that can be edited with an Intranet Web browser"
- ³ "GPO Seeks Tools to Harvest Overlooked Documents on the Web -- Government Computer News." http://gcn.com/articles/2005/01/03/gpo-seeks-tools-to-harvest-overlooked-documents-on-the-web.aspx?sc_lang=en (accessed 12/5/2009, 2009).
- ⁴ Goodwin, B., S. Campbell, J. Levy, Bounds, and Federal Web Managers Council. "Social Media and the Federal Government: Perceived and Real Barriers and Potential Solutions." (2008).
- ⁵ "Office of eDiplomacy." <http://www.state.gov/m/irm/ediplomacy/> (accessed 12/6/2009, 2009).
- ⁶ "Major Programs of the Office of eDiplomacy." <http://www.state.gov/m/irm/ediplomacy/c23840.htm#diplopedia> (accessed 12/6/2009, 2009).
- ⁷ "Knowledge Leadership at State." <http://www.state.gov/m/irm/ediplomacy/c23841.htm> (accessed 12/6/2009, 2009).
- ⁸ Westcott, Nicholas, Digital Diplomacy: The Impact of the Internet on International Relations (July 1, 2008). OII Working Paper No. 16, pg 3. Available at SSRN: <http://ssrn.com/abstract=1326476>
- ⁹ Westcott, Nicholas, Digital Diplomacy: The Impact of the Internet on International Relations (July 1, 2008). OII Working Paper No. 16, pg 9 Available at SSRN: <http://ssrn.com/abstract=1326476>
- ¹⁰ "Welcome to GreyNet, Your Grey Literature Network Service." <http://www.greynet.org/> (accessed 12/6/2009, 2009)
- ¹¹ "Proceedings:CB1 - Wikimania." <http://wikimania2006.wikimedia.org/wiki/Proceedings:CB1> (accessed 12/6/2009, 2009).

- ¹² Public Governance Committee. "Citizens at the Centre: Public Engagement for Better Policy and Services." *38th Session of the Public Governance Committee* (2008). 57
- ¹³ United States Department of State. *It Strategic Plan Fiscal Years 2006-2010: Empowering Diplomacy* 2007. pg 14
- ¹⁴ United States Department of State. *It Strategic Plan Fiscal Years 2006-2010: Empowering Diplomacy* 2007. pg 23
- ¹⁵ Westcott, Nicholas. Digital Diplomacy: The Impact of the Internet on International Relations (July 1, 2008). OII Working Paper No. 16. Available at SSRN: <http://ssrn.com/abstract=1326476>
- ¹⁶ Kelty, Christopher and George E. Marcus. "Open Source Experiments: What they show about the Analyst's Frustrations in Intelligence Communities." *Anthropology News* no. February (2007): 3-4., pg3
- ¹⁷ McConnel, Mike. "Collaborate, Cooperate, Integrate." *Unifying the Intelligence Profession (Pathfinder, May/June 2008)* (2008), 4.
- ¹⁸ Mintz, Daniel. "Government 2.0--Fact Or Fiction?" *Public Manager* 36, no. 4 (2007/2008): 22.
- ¹⁹ Goodwin, B., S. Campbell, J. Levy, Bounds, and Federal Web Managers Council. "Social Media and the Federal Government: Perceived and Real Barriers and Potential Solutions." (2008).pg2
- ²⁰ Goodwin, B., S. Campbell, J. Levy, Bounds, and Federal Web Managers Council. "Social Media and the Federal Government: Perceived and Real Barriers and Potential Solutions." (2008).pg 2
- ²¹ Goodwin, B., S. Campbell, J. Levy, Bounds, and Federal Web Managers Council. "Social Media and the Federal Government: Perceived and Real Barriers and Potential Solutions." (2008).pg 3
- ²² "DipNote : U.S. Department of State Official Blog." <http://blogs.state.gov/> (accessed 1/25/2010, 2010).
- ²³ "DipNote : U.S. Department of State Official Blog :: About." <http://blogs.state.gov/index.php/site/about> (accessed 1/25/2010, 2010).
- ²⁴ http://blogs.state.gov/index.php/entires/q_partnership_india/ (accessed 12/6/2009, 2009)
- ²⁵ "Dr. Shah's Statement | U.S. Department of State Blog." http://blogs.state.gov/index.php/entries/dr_shahs_statement/ (accessed 12/6/2009, 2009).
- ²⁶ Mahler, Julianne and Priscilla M. Regan. "Blogs as Public Forums for Agency Policymaking." *Issues in Governance Studies* no. 28 (2009): 1-14. pg7
- ²⁷ "Flickr: U.S. Department of State's Photostream." <http://www.flickr.com/photos/statephotos> (accessed 1/25/2010, 2010).
- ²⁸ <http://www.flickr.com/people/statephotos/>
- ²⁹ "Flickr: Explore U.S. Department of State's Photos on the Map." <http://www.flickr.com/photos/statephotos/map/> (accessed 1/25/2010, 2010).
- ³⁰ <http://www.facebook.com/usdos>
- ³¹ "U.S. Department of State | Facebook." <http://www.facebook.com/usdos#/usdos?v=info> (accessed 12/6/2009, 2009).
- ³² "U.S. Department of State - COP15 UN Climate Change Conference 2009 | Facebook." <http://www.facebook.com/usdos.cop15> (accessed 1/25/2010, 2010).
- ³³ "U.S. Department of State - COP15 UN Climate Change Conference 2009 | Facebook." <http://www.facebook.com/usdos.cop15> (accessed 12/6/2009, 2009).
- ³⁴ "U.S. Department of State - COP15 UN Climate Change Conference 2009 | Facebook." <http://www.facebook.com/usdos.cop15> (accessed 12/6/2009, 2009).
- ³⁵ "Careers in Foreign Affairs | Facebook." <http://www.facebook.com/group.php?gid=10328646356&ref=mf> (accessed 1/25/2010, 2010).
- ³⁶ "Careers in Foreign Affairs | Facebook." <http://www.facebook.com/group.php?gid=10328646356&ref=mf> (accessed 12/6/2009, 2009).
- ³⁷ "Civilian Response Corps of the United States of America | Facebook." <http://www.facebook.com/group.php?sid=a2bb63f7ed62196585568bc60eddfc6f&gid=35472323872> (accessed 1/25/2010, 2010).
- ³⁸ "Civilian Response Corps of the United States of America | Facebook." <http://www.facebook.com/group.php?sid=a2bb63f7ed62196585568bc60eddfc6f&gid=35472323872> (accessed 12/6/2009, 2009).
- ³⁹ "Bureau of Consular Affairs | Facebook." <http://www.facebook.com/pages/Bureau-of-Consular-Affairs/9229303148> (accessed 12/6/2009, 2009).
- ⁴⁰ "Diplomatic Security Service | Facebook." <http://www.facebook.com/group.php?gid=10493605641> (accessed 1/25/2010, 2010).
- ⁴¹ "Diplomatic Security Service | Facebook." <http://www.facebook.com/group.php?gid=10493605641> (accessed 12/6/2009, 2009).
- ⁴² "Department of State -- Office of Global Women's Issues | Facebook." <http://www.facebook.com/dos.sgwi> (accessed 1/25/2010, 2010).
- ⁴³ "Department of State -- Office of Global Women's Issues | Facebook." <http://www.facebook.com/dos.sgwi#/dos.sgwi?v=wall> (accessed 12/6/2009, 2009).
- ⁴⁴ "Twitter." <http://twitter.com/> (accessed 1/25/2010, 2010).
- ⁴⁵ "Twitter Blog: What's Happening?" <http://blog.twitter.com/2009/11/whats-happening.html> (accessed 12/6/2009, 2009).
- ⁴⁶ "Travel - State Dept (TravelGov) on Twitter." <http://twitter.com/TravelGov> (accessed 1/25/2010, 2010). \
- ⁴⁷ "Travel - State Dept (TravelGov) on Twitter." <http://twitter.com/TravelGov> (accessed 12/6/2009, 2009).
- ⁴⁸ "YouTube - Statevideo's Channel." <http://www.youtube.com/user/statevideo> (accessed 1/25/2010, 2010).
- ⁴⁹ "YouTube - Statevideo's Channel." <http://www.youtube.com/user/statevideo> (accessed 12/6/2009, 2009).
- ⁵⁰ <http://www.youtube.com/watch?v=oSGFP1zvjmcc>
- ⁵¹ "The Constitution of the United States" Article I, Section 5, Clause 3.
- ⁵² Buckley, C. W. . *U.S Government Documents*: Graduate School of Library Science. University of Illinois at Urbana-Champaign., 1963.



IIa | Information International Associates

Creating value from the world's information

- Scientific and Technical Information Management
- Open Source Intelligence Research
- Information Technology Services
- Library and Information Center Management
- Records and Technical Documentation Management
- Education and Training Program Administration



www.iiaweb.com

Information International Associates, Inc (IIa)
P.O. Box 4219
1055 Commerce Park Drive, Suite 110
Oak Ridge, TN 37830
Phone: 865.481.0388 Fax: 865.481.0390
E-mail: inquiry@iiaweb.com

The Grey System for Monitoring Self-Funded Research *

Aleksandr V. Starovoitov, Yuri M. Bogdanov, Aleksandr M. Bastrykin, and
Leonid P. Pavlov (Russia)

Abstract

To eliminate the defects in research monitoring in Russia a special Government decision was issued in 2006 with the idea of creating a system for self-funded research projects registration. The system was designed in the years of 2007 – 2008 and now put into operation as an integral part of the federal scientific and technical information grey literature system. Based on the output information from the system the annual summary report for the Joint Interministerial Commission specified in the Government Decision is prepared. A general outline of the system, the description of its input and output document forms and database structure are given in the paper.

Introduction

Today's global economy is often described as one in transition to a "knowledge economy", or "knowledge-based economy" which is being developed in the environment of a "knowledge society" as an extension of an "information society". The theoretical grounds for this kind of treatment as well as the very terms "information society", "knowledge society" and "knowledge economy" were put into scientific circulation as long ago as in the late sixties last century by the famous American economist and the founder of the modern management theory Peter Drucker in his book "The Age of Discontinuity" that was republished many times since then [1].

Now that the Drucker's forecast came true the decisive role of human capital in the form of knowledge and education is universally recognized as a powerful productive asset and the source of innovative solutions. In Russia of today it is well understood that the way from the backward raw materials economy to the modern knowledge economy lies through innovations generated by scientific research and applied in industry. Some burdens of the Soviet science strategies still haunt the Russian science even now and about 80% of financial support for science still comes from the state [2] while Russian business is reluctant to invest in scientific research. Also, it was typical for the Soviet science to carry out research at the widest front possible disregarding the practical results if any and this approach can not be followed today. So, new concepts are required in the state scientific policy.

Administrative measures

It is evident that the immediate aim of forming a knowledge society and an innovation economy suggests an adequate funding of scientific research and development. There has been a growth in scientific research state budgeting every recent fiscal year (except this year of 2009 when because of the world economic crisis the state budget of Russia was sequestered including the item of science expenditure). So, the federal budget science expenditure in percentage to the total budget expenditure grew from 1,69% in 2000 to 2,27% in 2006 [3]. In absolute figures the growth of federal budget science expenditure is as follows: 72,4 billion rubles in 2006, 89 billion rubles in 2007 and 118.4 billion rubles in 2008 [4].

The figures of budget assignments to science being important it is no less important to optimize the distribution of financial means among the fields of science. To do so priorities in scientific research and critical subjects should be determined. Since there are about one hundred budget recipients in Russia - federal ministries and agencies that have their share in the total budget science expenditure to finance their scientific research and development projects - Joint Interministerial Commission was formed as a decision-making body for research policy and with the aim of working out or updating the list of science and technology priority development directions and the list of critical technologies in the Russian Federation at least once in four years.

Now there are totally 8 priority directions, for example bio-systems, energy and energy saving, information and telecommunication systems, nano-systems and materials, and 34 critical technologies like bioinformation, software design, hydrogen

* First published in the GL11 Conference Proceedings, March 2010

energy, distributed computing, nuclear energy and fuel, cell technologies, atmo- and hydrosphere monitoring, refuse utilization, etc. The lists prepared by the Commission on the basis of different means of science monitoring and statistics are approved by President of the Russian Federation (the latest approval dated May 21, 2006).

Collecting and monitoring state-funded R&D

One of those means is the federal scientific and technical information system for grey literature that supports monitoring (both in financial and subject respect) of the state funded scientific research and development activities covering extensively all the territory of the Russian Federation [5]. The system collects and controls scientific and technical reports and dissertations concerned basically all scientific subjects ranging from mathematics, physics, electronics and engineering through to social sciences and the humanities.

The legal ground for the operation of the system is the Federal Law "On the obligatory copy of documents" that makes it mandatory for all the scientific research executors (scientific and higher school institutions, industrial organizations) to be the input documents suppliers to the system in the form of full-text research and development reports accompanied by registration and information cards with the abstracts and metadata of the documents. Also, data on contracts and funding arrive from federal ministries and agencies and dissertations come from dissertation councils and individual researchers.

Since 2004 the full-text documents have been scanned and converted into a digital representation to form a full-text digital repository (before 2004 the full-text reports and dissertations were microfilmed) and since 1982 the registration and information cards with the abstracts and bibliographic descriptions of the documents have been keyed-in into computer to form the corresponding retrospective database serving an electronic catalogue to the repository and a means of quick and easy online search and access to the information. Each document has a unique inventory number both in the database and in the repository. Now there are about 3 million documents in the database.

The users of the system are able to conduct online subject search and order copies of documents both via the Internet and in the reading room. Most individual users are researchers, experts, lecturers and tutors, under- and post-graduate students. Among collective users are different kinds of organizations – scientific research institutes, universities, industrial and business community, Higher Dissertation Commission, state executive and controlling authorities (ministries, agencies, audit and fiscal bodies, law-courts). Official users usually require statistical, financial and generalizing information on science.

Thus, the existing system for grey scientific and technical information sources is the most complete and reliable means of monitoring and controlling the situation in state-funded scientific research and development activities. The system's collection is an indispensable source for government agencies with an interest in the latest Russian contributions to science and technology.

Self-funded research monitoring Decision

No matter how much money is given to science from the state budget it can never be the only and sufficient financial source for research and development. So, the diversification of funding is inevitable and there is a growing trend that more and more scientific projects are being funded from research organizations' own financial resources.

Those organizations are commercial state-owned ones functioning in the forms of federal state unitary enterprises and open joint-stock companies with the state shareholding. Their self-funded research projects are not covered by the federal grey literature system and hence were out of centralized monitoring and were not taken into account by the Joint Interministerial Commission when preparing proposals for updating the lists of priority directions and critical technologies.

To eliminate the defects in research monitoring a special Government Decision was issued on November 4, 2006 (No. 645) with the idea of creating a system for self-funded research projects registration. The greatest difficulty with such systems of all-Russian scale is not its computer network and software realization (though not a

simple thing in itself) but its, so to say, organizational technology - the ensuring that all the commercial organizations involved provide their reports to the system timely, complete and correct even if they are supposed to do so in accordance with the Government Decision. Unfortunately, the sad experience of the latest decades exposed a very low executive discipline of scientific and scholar institutions. So, Government Decision No. 645 not only stipulated the idea but also defined the mechanisms of its realization.

It is well known that financial reports are presented by all the organizations most unquestioningly and without demur. The Decision obliged the top managers of the federal state unitary enterprises and open joint-stock companies with the state share-holding to include in their annual financial report presented to their higher ministry or agency the information on their self-funded research and development projects and to submit this information to the Federal Agency on Science and Innovations responsible for the scientific activities assessment and monitoring in the Russian Federation. The Decision also obliged the Federal Agency on Science and Innovations to generalize the received information and issue the Annual Summary Report on self-funded research intended for the Joint Interministerial Commission so that it could take into consideration the data from the Report when working out the lists of priority directions in science and critical technologies.

In accordance with the Decision the information on self-funded research should be submitted in an approved unified form as an annex enclosed in the organization's annual financial report. The approved blank form is added to the Decision's text. The form's fields of data are important because their filling determines the information value of the document. Of course, the form is in the Russian language and we will consider its translation into English.

The heading of the form is as following: "Civil purpose self-funded scientific R&D projects of the federal state unitary enterprise (or the open joint-stock company with the state share-holding)" – the specific organization's name and report year are to be filled-in. Next come the following names of fields to be filled in:

- city of the organization
- postal address and phone number
- higher federal authority body (ministry, agency, etc.)
- branch of activity
- main kind of activity.

After those attributes defining the organization the fields follow that characterize the content of research, its time limits and self-funding:

- subject(s) of scientific R&D (project-title)
- priority direction, critical technology
- term of carrying out (years)
- volume of financing for the fiscal year (in thousand rubles)
- basic results (in increasing way, for the fiscal year included).

The content fields are filled-in in a free way. The filled-in form is signed by the head of organization, stamped and directed to the Federal Agency on Science and Innovations. The form is to be submitted both in printed (on paper) and in digital (Microsoft Word) shape.

The System

To carry into effect the Government Decision a system based on computer network and database technologies was designed in the years of 2007 – 2008 and now put into operation as an integral part of the federal scientific and technical information grey literature system. The self-funded research monitoring system is evidently grey because its input form described above and output Annual Summary Report are typically grey documents.

The state contract for the system's design and introduction was won by the Centre of Information Technologies and Systems of Executive State Authorities (CITIS) that is now merged with the Scientific and Technical Information Centre of Russia (VNTIC) being the leading centre for grey literature in Russia. The system's architecture and software was developed by the specialists of CITIS and VNTIC.

The heart of the system is the database on self-funded research with the filled-in annex forms. It was not possible to completely avoid the paper forms since the annex

was enclosed in the annual financial report submitted on paper. Nevertheless, an online mode of filling the annex form is provided for the authors of self-funded research projects who are able to address the CITIS site on the Internet (www.citis.ru), click "the annex form online filling-in" and have the form on the screen of their computer. There are many conveniences supporting the online filling-in such as the formal verification of numerical fields, the enclosed lists of priority directions and critical technologies and the list of correct names of the organizations that were previously registered in the system. The user just has to click the name instead of keying it in.

On completing the online filling-in of the annex form the user directs the document via the Internet straight to CITIS where it is entered into an intermediate (technological) database. At the same time the user prints out the completed form, this paper form (signed by head of the organization and stamped) is added to the annual financial report and submitted to the Federal Agency on Science and Innovations wherefrom it arrives in CITIS.

The documents in the technological database and the corresponding documents arrived on paper are verified, after verification are entered into the database on self-funded research and each assigned a unique registration number. This main database of the system is provided with a well-developed search and retrieval means and all the fields of the document are searchable. The database management system or rather information search and retrieval system is IRBIS designed by a group of Russian scientists and programmers under the supervision of Professor Nicolai Maximov [6] and now recognized as a de-facto standard DBMS in the sphere of scientific and technical information in Russia.

Now there is a two-year retrospective database (with the report documents of 2007 and 2008 years – totally about 1 thousand documents), next year (2010) the documents of 2009 will be entered and so on. Thus, the system ensures the registration of report documents on self-funded research, their permanent storage in the database format and both quantitative and qualitative analysis of self-funded research in Russia prepared in the form of the Annual Summary Report.

Since the field "Basic results" in the annex form often is filled-in by the report authors in rather a short way it is not informative at times. Sometimes the same is true for the field "Subjects of scientific R&D" when the title of a project is rather general and not specific. In these cases it is difficult to analyze the research results qualitatively. In order to enhance the system's potentialities a link between its database and the Federal registration and information cards database (the state-funded research database) was envisaged. The documents of the latter database – the information cards – are more detailed than the annex form and contain the abstracts of full-text reports.

To make the link work it is necessary to persuade the authors of self-funded research projects that they should also register their projects in the Federal grey literature system and fill-in the information card so that it could be entered into the Federal database. In case of the lack of information in the annex form for a detailed analysis it would be possible to address the corresponding information card. A circular letter with such a recommendation was directed to all the federal state unitary enterprises and open joint-stock companies with the state share-holding. Unfortunately, only three organizations followed the recommendation and registered their 22 projects in both the databases.

Annual Summary Report

The main output document of the system is the Annual Summary Report on self-funded research intended for the Joint Interministerial Commission. The Report consists of the textual analysis and the illustrative materials in the form of tables and pie charts. So far two Reports were prepared by means of the system – in 2008 (documents of 2007) and in 2009 (documents of 2008).

The major items of the Report contents are as follows:

- table of government bodies showing the amount of their dependent enterprises and companies submitted the self-funded R&D information

- distribution of organizations' financial means for self-funded R&D by the federal government bodies
- self-funded R&D financing distribution by priority directions and critical technologies
- the enterprises and companies spent more than 50 million rubles on self-funded R&D
- comparative tables of data for the reported and previous year
- generalized data and its analysis
- conclusions and recommendations.

In 2007 and 2008 the total average amount of financial means spent by enterprises and companies on self-funded research per year is more than 1 billion rubles. In 2008 the amount of financial means spent on self-funded research concerned the priority directions and critical technologies increased nearly three times as compared to the year of 2007. In 2008 totally 56 organizations (39 enterprises and 17 companies) funded 455 research and development projects (the enterprises – 297 projects and the companies – 158 projects) from their own financial resources. Three enterprises spent more than 50 million rubles on self-funded research.

As an example let us consider a pie chart reflecting the self-funded research distribution by priority directions of scientific and technological development (in percentage of spent money):

- rational wildlife management – 41.14 %
- energy and energy saving – 24,92 %
- live systems – 19,20 %
- nanosystems and materials – 10,28 %
- transport, aircraft and space systems - 2.10 %
- security and antiterrorism - 1,46 %
- information and telecommunication systems - 0,90 %.

Concluding Remarks

Now that the system is put into operation it may be concluded that self-funded research is no longer out of monitoring and control. The information obtained by means of the system allows:

- to monitor the situation in the sphere of self-funded R&D all over Russia
- to update the lists of priority directions and critical technologies
- to improve the distribution of financial means for scientific R&D both federal budgeted and self-funded
- to reduce the unjustified duplication and overlapping of R&D projects.

References

1. Drucker, Peter F. The Age of Discontinuity: Guidelines to Our Changing Society. – Rutgers Univ., N.J.: Transaction Publishers, 1992. – 420 pp. - ISBN-10: 1560006188. ISBN-13: 978-1560006183.
2. Kuznetsov, Yuri. Federal budget funding of civil science in Russia. / Otechestvennye zapiski. – No.7, 2002 (In Russian).
3. Krichevsky, Nikita. Crisis, gentlemen! <http://www.apn.ru/publications/article20966.htm> (In Russian). Visited 22.09.2009.
4. Ministry of Finance of the Russian Federation official site <http://www1.minfin.ru/ru/budget> (In Russian). Visited 22.09.2009.
5. Pavlov, L.P. The State and Development of the Russian Grey Literature Collection and Dissemination Centre. – Interlending and Document Supply. - MCB Univ. Press. 1998, vol. 26, No. 4. Pp. 168-170.
6. Maximov, N.V. et al. Information Systems. – Moscow. – FORUM: INFRA-M, 2007. – 496 pp. (in Russian).

Largest collection of e-Journals in Japan

Current issues

J-STAGE will assist
your research
by providing you
with the latest world class
scientific information.

J-STAGE

- All abstracts **FREE** (90% with English)
- More than 570 **PEER-REVIEWED** journals (70% **FREE**)
- **FULL TEXT SEARCH**
- **ADVANCE** online publication
- Personalized management tool – 'My J-STAGE'

www.jstage.jst.go.jp

Back issues

Journal@rchive is
a web based electronic
archive initiative offering
the most highly-valued
academic journals in Japan.

Journal@rchive.

- **Over 750,000** articles published by Japanese societies
- **FREE** from 1st issues
- The oldest goes **back to 1877**
- **FULL TEXT SEARCH**

www.journalarchive.jst.go.jp



Japan Science and Technology Agency
ELECTRONIC JOURNALS DIVISION
DEPARTMENT OF ADVANCED DATABASES

contact@jstage.jst.go.jp

An Infrastructure for Managing EC Funded Research Output : The OpenAIRE Project

Paolo Manghi (Italy), Natalia Manola (Greece), Wolfram Horstmann (Germany), and
Dale Peters (South Africa)

Abstract

The OpenAIRE project aims at supporting the diffusion and adoption of the European Commission Open Access mandates among researchers in Europe. It will deliver an organizational/technological infrastructure "supporting mechanisms for the identification, deposition, access, and monitoring of FP7 and ERC funded articles". The organizational infrastructure will provide a European Helpdesk jointly operated by a European network of National correspondents. The technological infrastructure will be based on state-of-the-art software services of the D-NET Software Toolkit developed within the DRIVER and DRIVER-II projects and the Invenio digital repository software developed at CERN.

Scenario and Motivations

"All research builds on former work, and depends on scientists' possibilities to access and share scientific information. The advent of the internet and electronic publishing, have resulted in unprecedented possibilities for the dissemination and exchange of information. '**Open Access**', defined as free access over the internet, aims to improve and promote the dissemination of knowledge, thereby improving the efficiency of scientific discovery and maximising return on investment in R&D by public research funding bodies."¹

Although simple in conception, unrestricted availability of research publications and scientific data is still far from reality; the implementation of policies that promote Open Access to these most important products of research has proved to be very challenging. The "Handbook" on "Open Access – Opportunities and Challenges"² **Fout!** **Verwijzingsbron niet gevonden.** provides a comprehensive overview of the complexities of all relevant policy, organisational, social and academic cultural, technical, financial, and legal issues. Many initiatives and organisations (SPARC Europe, The Knowledge Exchange³ and The European Universities Association's expert Working Group on Open Access⁴) in EU member states and at trans-national level have achieved remarkable progress in the Open Access and Digital Repositories areas over the last years, proving that these complexities can be overcome.

European Commission Open Access Mandates

In August 2008, the European Commission launched an open access pilot in FP7⁵ under which, grant recipients in seven areas (energy, environment, health, parts of information and communication technologies, research infrastructures, science in society, and social sciences and humanities) will be required to deposit peer reviewed research articles or final manuscripts resulting from their FP7 projects into an online repository, and make their best efforts to ensure open access to these articles within either 6 or 12 months after publication⁶.

In addition, the ERC Scientific Council's Statement on Open Access of December 2006⁷ and the subsequent December 2007 Guidelines require that all peer-reviewed publications from ERC-funded research projects be deposited on publication into an appropriate research repository where available, such as PubMed Central, ArXiv or an institutional repository, and subsequently made Open Access within 6 months of publication. ERC also considers essential that primary data are deposited to the relevant databases as soon as possible, preferably immediately after publication and in any case not later than 6 months after the date of publication.

OpenAIRE goals and strategy

OpenAIRE (Open Access Infrastructure for Research in Europe) is a three year project that will realize the EC Open Access pilot and assist in the ERC OA guidelines providing an electronic infrastructure and supporting mechanisms for the identification, deposition, access, and monitoring of FP7 and ERC funded articles. Thematically, it will focus on peer-reviewed publications in at least the seven disciplines highlighted in the OA Pilot and on research datasets. Geographically, it will have a definitive "European footprint" by covering the European Union in its entirety, engaging people and scientific repositories in almost all 27 member states.

All deposited articles that result from EU-funded research projects will be freely accessible through the www.openaire.eu portal, which will also provide tools and online support for assisting researchers to publish (deposit) EC funded publications in two ways: (i) by establishing and operating a special repository for articles that can be stored neither in institutional nor in subject-based/thematic repositories, and (ii) by addressing the "depositing only once" issue by developing ways to mediate between the central portal (i.e. OpenAIRE) and the local repositories. Finally, it will provide tools to funders for measuring and evaluating the EC funded research output and the effect of the mandates.

To achieve its goals OpenAIRE concentrates and addresses the following objectives:

OBJECTIVE 1:

Building Support Structures for Researchers in Depositing FP7 Research Publications

The FP7 OA Pilot mandate has been an indispensable step towards free access to European research results. Experiences with other OA mandates, however, show that acceptance and broad take-up by the scientific community critically depends on accompanying support mechanisms, as any activity beyond the actual research and publishing process is considered by researchers as administrative burden. OpenAIRE will provide support structures assisting them in quick and efficient article depositions, thereby ensuring that a critical mass of articles will be deposited.

OBJECTIVE 2:

Establishment and Operation of the OpenAIRE e-Infrastructure for Peer-Reviewed Articles and Other forms of Scientific Results

OpenAIRE will realize and maintain a system based on state-of-the-art software services of the D-NET software toolkit⁸ developed within the DRIVER⁹ projects and the Invenio digital repository software developed at CERN¹⁰. These will be further enhanced and complemented with services developed within OpenAIRE to show links of publications to the EC funded projects. Furthermore, monitoring tools and statistics services will infer relevant information and statistics on FP7 and ERC funded research from article, research data, project metadata and relationships between them.

OBJECTIVE 3:

Exploration of and experimentation with Scientific Data Management Services

OpenAIRE will work with several subject communities to explore the requirements, practices, incentives, workflows, data models, and technologies to deposit, access, and otherwise manipulate *research datasets* of various forms in combination with research publications. Feasibility studies will be produced to explore all relevant complex processes & structures and show the benefit for researchers in both depositing and re-using these combined information resource packages. OpenAIRE will investigate supporting potentially all scenarios mentioned earlier about peer-reviewed articles and will choose a subset of those in the actual studies.

OBJECTIVE 4:

Sustainability of the OpenAIRE e-Infrastructure

Sustainability of OpenAIRE will only be ensured if there are dynamic developments, with new repositories joining the e-Infrastructures, new documents being uploaded, and new researchers making use of the services offered. OpenAIRE aims at marshalling in a user-friendly way, existing strategies to promote the Open Access principle and the use of its e-Infrastructure, and to demonstrate the advantages of integrated efforts with research funding organizations and information infrastructures thus obtained for scientific and scholarly communities.

The project's support for the establishment of OA offices in all EU member states is particularly important as several European countries have not put Open Access or repositories on their national agendas¹¹. This will have significant impact on a broad uptake of Open Access at a pan-European scale, thus ensuring the livelihood of the OpenAIRE tools and services. Moreover, OpenAIRE's synergies with other relevant initiatives and its close link to COAR¹² - Confederation of Open Access Repositories - will make OpenAIRE's outcomes visible to a broader, international audience.

Paper focus and outline

Section 0 provides a synopsis of OpenAIRE's *organizational infrastructure* offering services to support researchers over the 27 partner European Countries towards meeting the demands of the FP7 OA mandate (Objective 1). Section 0 presents a more detailed overview of OpenAIRE's *technical infrastructure* offering applications for the

population, access and manipulation of a *European Research Information Space of OA* publications and their relationships with the EU administrative project data (Objective 2). Section 0 provides an outline of how OpenAIRE plans to investigate and experiment on data models and functionalities for the future extension of the technical infrastructure with applications for managing research data and how to combine it with related publications in the Information Space (Objective 3). Finally, Section 0 comments on future issues and possibilities.

Organizational infrastructure

Authors and researchers in general must be informed on the various aspects of the depositing process related to the OA mandates: they must be aware of issues ranging from different depositing process scenarios (e.g. self or mediated) and guidelines, publishing and licensing models, copyright/licensing policies (e.g. Romeo¹³ database on copyright policies), the difference between author versions vs. publisher versions, to policies for requesting authorization to self-archive after embargo. Moreover, researchers need to fulfil the OA terms by operating on the local level, in the context of their affiliation with local universities and research institutes. One of the focal points of OpenAIRE is to provide all the necessary support through a series of tasks carried out in 27 European countries

Its main tasks will be (i) to develop an European Helpdesk System, comprising a European Centre and national Open Access liaison offices, and (ii) to liaise with other Open Access and repository related activities in Europe in order to cascade information to local decision makers in the national context of researcher affiliation to research organizations, or discipline related research activities.

Pan European Helpdesk System

The provision of a continuous operational service to facilitate deposit is critical to the support envisaged for researchers. The Helpdesk functionality is not limited to standard online query submission and fast response times, but acts as a network providing supporting measures that covers all EU member states, and offers seamless information on issues ranging from the mapping of a point of queries to the location of the nearest relevant repository, localised information on intellectual property rights and deposit guidelines. In particular, the European Helpdesk System will offer:

- FAQ's, selected list of relevant references (repositories, sites, services, organizations, and experts), as well as relevant materials for incremental distribution to national liaison offices.
- Toolkits for researchers to guide them through the deposition process (advocacy material and information on copyright/licensing policies of selected journals); toolkits for research institutions on OA policy development and compliance with the EC mandates.

The organizational structure of the Helpdesk (Figure 1) is moderated and administered centrally, with distributed responsibility among experienced partners and a network of national OA Liaison Offices.



Figure 1. OpenAIRE Organizational Structure.

Liaison with OA and repository related activities in Europe

The FP7 OA pilot is one of many parts of the developing landscape of Open Access activities in Europe. The overlapping affiliation of OpenAIRE participating organizations with major ongoing initiatives and its strong link to COAR, ensures collaboration with, and inclusion of all relevant stakeholders. The envisioned networked liaison scheme includes building collaborative initiatives with publishers through the PEER project¹⁴, with organizations developing and maintaining Current Research Information Systems (CRIS), and with major European OA organizations (SPARC Europe and the OA Working Group of the European University Association).

Technological infrastructure

The OpenAIRE *portal* (www.openaire.eu) is the gateway to all underlying applications and services provided by the technical infrastructure, where researchers or organizations come to deposit their research outcome, search and access information on related publications or European projects, receive information on OA policies, licenses and copyright issues, or even discover a link to an institutional or subject repository matching their location or discipline. The targeted user typologies are:

Authors interested in depositing their publications into the infrastructure associating it with the FP7 project information;

Researchers and general public interested in accessing the Information Space of all OA FP7 publications through friendly web interfaces and a variety of functionalities;

Commission and organizational funders studying and assessing the effectiveness of their policies on research directives as well as publication models;

Repository community interested in becoming an integral part of the OpenAIRE infrastructure integrating deposition and publication referral processes;

E-Science applications operated by third-party organizations, interested in accessing/retrieving content from the Information Space.

Concrete scenarios have been defined for the construction of an integrated European research information system. Although publication deposition and their relation to projects is the central use case, other real-case scenarios, equally important, help at formulating the system components.

Scenario 1:

Depositing a publication in the European Research Information Space

The system will support two high-level deposition patterns.

Autonomous deposit: authors deposit autonomously eligible FP7 and ERC-funded results in existing repositories without direct relation to OpenAIRE.

OpenAIRE mediated deposit: Authors come to OpenAIRE and expect guidance for deposition towards existing repositories, or deposit directly into the OpenAIRE provided "orphan repository"¹⁵.

Scenario 2:

Accessing the European Research Information Space

End-users want to retrieve information on publications and project data by: (i) Searching/browsing the content, e.g. breakdown and display by research, projects, people, funds, publications, subjects, etc., (ii) Converting publications on-demand into other types of MIME formats; (iii) Receiving notifications on events related to their expertise and type of work, such as: impact (derived from overall usage statistics) on their publications or projects; new projects or publications appearing within their scientific research area; publications released from specific projects or more general research areas or on usage statistics (calculated by thresholds) on specific research areas/projects.

Scenario 3:

Monitoring and assessing impact through usage statistics

Users want to measure the impact of publications and projects through usage statistics (obtained from the orphan repository and all other participating repositories) and their combination of various quantitative web indicators involving external link references and public search engines data.

Scenario 4:

Third-party services accessing the Information Space

Since OpenAIRE is envisioned to be a part of the European e-infrastructure ecology, its content may be re-used by third-party services, e.g. aggregators, distributed search engines, e-Science text-mining applications. Such services and e-infrastructures (e.g. Europeana¹⁶, DRIVER) will be able to retrieve the OpenAIRE rich content through a variety of protocols.

Architecture and technologies

The OpenAIRE technical e-infrastructure provides a system capable of supporting the functionalities identified by the scenarios above, integrated with the tools of the European Helpdesk to present end-users with one, integrated point of entry. As illustrated in Figure 2, the system is structured into four main functional areas: *presentation area and deposition area*, dealing with users accessing and populating the Information Space (system front-end, scenarios 1 and 2); *storage area and mediation area*, dealing with Information Space content management, including storage and access through standard protocols from third party-services (system back-end, scenarios 3 and 4).

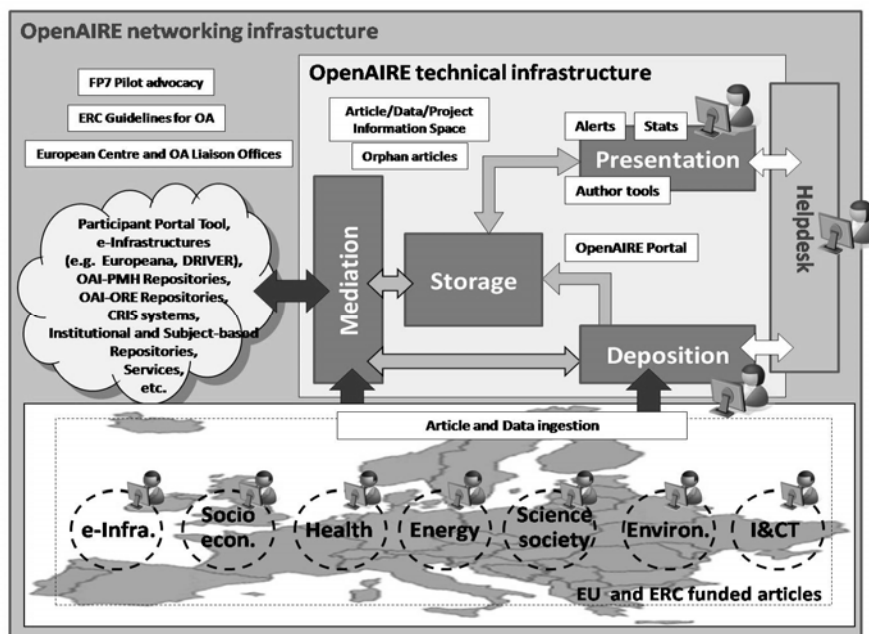


Figure 2 – OpenAIRE system conceptual architecture.

The system consists of a number of service-oriented applications, “glued” together with the infrastructural application framework offered by the *D-NET Software Kit*^{17,18}, a component oriented approach based on a *Service Oriented Architecture* (SOA)^{19 200}, developed during the DRIVER projects. D-NET service applications are *open*: extendable with new service typologies, i.e. new functionalities; *interoperable*: services can encapsulate heterogeneous technologies; *flexible in architecture*: services running on the same site behave like components of a centralized system.

The D-NET Software Kit services are logically organized in three architectural layers: *Functionality Layer Service Kit*. Includes the services needed to form portals: configurable user interfaces (portals), recommendation systems, user and community profiling, record collections management etc. Appropriately combined, such services can form a variety of community applications, which can then be applied to operate over an information space.

Data Layer Service Kit. Includes the services needed to form aggregation system applications, whose functionalities feature harvesting, cleaning, storing, indexing, searching the metadata records harvested from external OAI-PMH Repositories.

Enabling Layer Service Kit. Includes the Services supporting the application framework. These provide “gluing” functionalities such as service registration, discovery, orchestration, user authentication and authorization, subscription and notification.

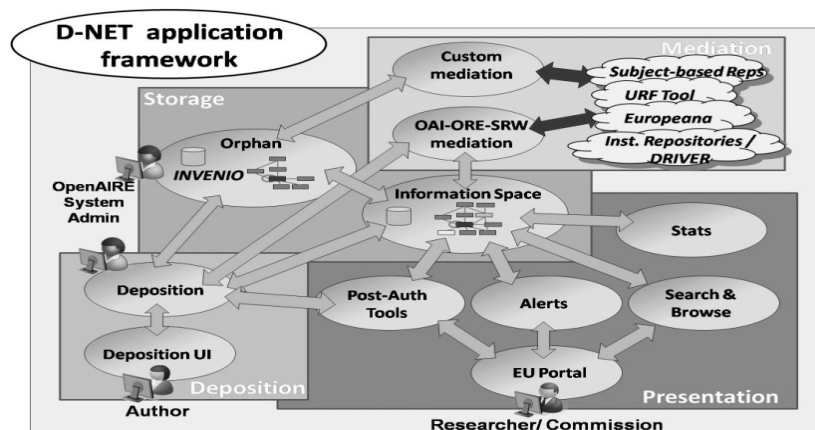


Figure 3. OpenAIRE services running on D-NET framework.

The choice of D-NET for OpenAIRE was motivated by four main factors: (i) its “gluing features” and “openness”, enabling for technological interoperability and future expansion; (ii) its flexible architecture, supporting both centralized and distributed

scenarios; (iii) its proved success in other projects (DRIVER and EFG²¹) for analogous services; (iv) many of the existing services used in DRIVER (harvesting, aggregating, storing, searching) will be adapted and transferred to OpenAIRE.

The design and development of the system is organized in three main steps described in detail in the following sections: Setting up the OpenAIRE Information Space and Orphan Repository, Populating the OpenAIRE Information Space, Accessing and elaborating the OpenAIRE Information Space.

Setting up the OpenAIRE Information Space and Orphan Repository

OpenAIRE supports an Information Space containing publication and administrative EU funded projects data. The system operates under a common data model (the *OpenAIRE data model*), formally describing the structure of the content: how publication and project metadata are structured and associated with each other, extended to include research data descriptions (data sets) and usage statistics data with the corresponding relationships to projects and publications. The Information Space Service offers efficient storage, access and management to the linked OpenAIRE objects, so as to support depositing (Scenario 1) and accessing (Scenarios 2,3,4) of publication-project metadata, as well as usage statistics concerning all FP7 OA publications.

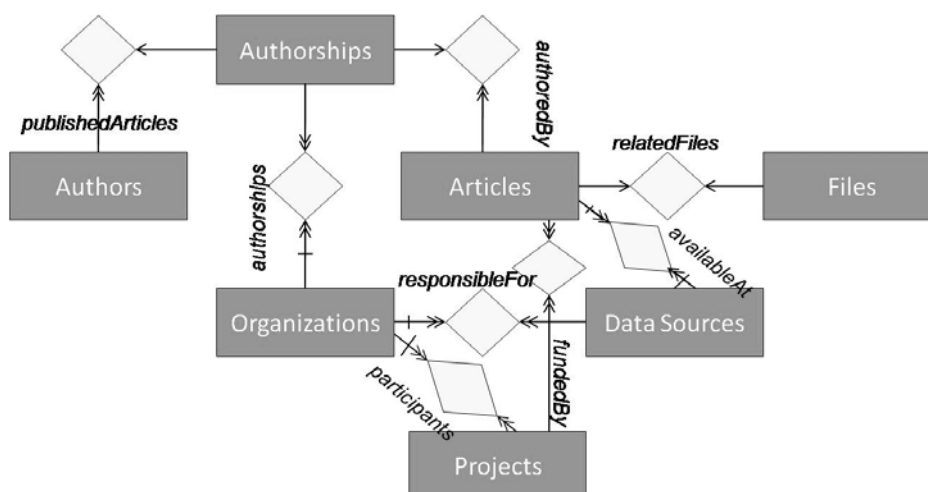


Figure 4. Indicative example of the OpenAIRE data model.

Populating the OpenAIRE Information Space

The OpenAIRE system enables deposition and ingestion mechanisms (scenario 1) of publication and project metadata. Authors accessing the OpenAIRE portal to deposit their publications are driven through one of the following scenarios:

Manual deposition from authors, where they (i) either deposit the full payload (e.g. pdf, data set) into the OpenAIRE 'orphan' repository, or (ii) deposit the "information about the publication" (i.e. bibliographic metadata) while the publication is being, or has been deposited in an OA repository. For (i), the Deposition Service is closely linked to the *orphan* repository, integrating its UI, user identity management and storage, while for (ii), the Deposition Service provides visual tools enabling authors to *identify* their publication (e.g. provide repository name, publication DOI), and then to complete the ingestion by linking it to the *authoritative* EU project metadata provided by OpenAIRE.

Ingestion of publications from repositories, in which case the most important task is to identify and automatically import the publications with the EC-funded footprint, and assign these to the corresponding projects. Technical solutions require the synergy of OpenAIRE and the repository community in order to define and apply common guidelines for representing and exposing EC-funded publications metadata, i.e. bibliographic and project information, and ultimately implement solutions for notifying OpenAIRE when new publications have been deposited in a repository. Where common guidelines cannot be fully applied, manual and intellectual curation activities verify and fix the operation of this automatic process.

Ingestion of authoritative EU funded project data from the Commission's Participant Portal. The Participant Portal is the gateway to the database maintained by the European Commission to store all its funding related activities (i.e. proposals, projects, institutions, project officers and coordinators, etc.). OpenAIRE, with its extended

services, retrieves all the *authoritative* project information and ingests it into its Information Space according to the OpenAIRE data model.

Accessing and elaborating the OpenAIRE Information Space

The OpenAIRE system provides interactive access (scenarios 2, 3) to publication and project metadata in the Information Space. Through the OpenAIRE web portal end-users are able to search, browse, access the content. A prototype framework for format conversion services is provided in order to enable users to access documents according to their format of preference.

Through the development of usage statistics services OpenAIRE exploits usage metrics as a supplement to conventional citation analysis (usage metrics are a promising new field of research indicators and citation analysis is too slow to yield results for publications resulting from FP7 and ERC funding within the lifetime of the project). Based on recent or emerging standards and protocols (PIRUS2²², SUSHI²³) different metrics for usage and aggregation activities from various distributed data sources are developed. Usage data will be harvested and aggregated from local institutional or subject-based repositories and the OpenAIRE portal itself in the form of logging usage events (hits, downloads etc.) with server sided tools (e.g. apache Log-files). In combination with 'webometric' data ("web-citations", backLinks)²⁴ and regular citations OpenAIRE intends to provide impact monitoring. Presentation of usage data is planned to take two different forms: (i) an article that is represented through the OpenAIRE portal has additional information on usage in the form of "downloaded *n* times" (aggregated usage information from local repositories), "accessed through OpenAIRE *n* times", "linked from sites *a*, *b*, *c* ..." and "cited by *x*, *y*, *z*..."; (ii) comparative reports (written or software based) as a monitoring activity.

Finally, third-party providers may access the information space in a non-interactive way (scenario 4). This involves organizations willing to take advantage of the enriched research information (e.g. links between publication and project data) and its derived results (e.g. usage statistics). Standard protocols, namely OAI-PMH and SRW are delivered, but OpenAIRE also intends to provide customized APIs for the interaction with organizations that are not compliant with the above protocols.

Research data and publications

One of the key objectives of OpenAIRE is to perform studies on subject specific requirements for primary and secondary research data in the seven research areas of the FP7 pilot. The expected outcome of such investigation is a book that describes the data models that best capture structure and semantics of the relation between publications and research data in each of the areas to be studied. In particular, the project focuses the analysis on four out of the seven research areas of the Pilot: 'health' as exemplified by basic research in the life sciences, 'environment' as exemplified by research on climate and agriculture, ICT (Challenge 2: Cognitive Systems, Interaction Robotics), and socio-economic sciences and humanities. The input from those areas is provided by five project partners who are heavily involved in other major European projects:

- *Health (Life Sciences)*: The European Bioinformatics Institute²⁵ (EBI) as part of the European Molecular Biology Laboratory (EMBL)²⁶ connecting, among other high-impact services and projects in basic life-science research and health.
- *Environment*: The World Data Center for Climate²⁷ (Max-Planck Institute for Meteorology²⁸) and the Consultative Group on International Agricultural Research (CGIAR²⁹) researching the sustainable use and conservation of agricultural biodiversity.
- *Information & Communication Technology (Cognitive systems, Interaction, Robotics)*: The Center of Excellence: Cognitive Interaction Technology³⁰ (CITEC) at Bielefeld University.
- *Socio-economic Sciences and Humanities*: Data Archiving and Networked Services³¹ (DANS) connecting to European research infrastructures.

Conclusions and Future work

OpenAIRE and the FP7 pilot have started out an important and long term vision for EC funded research. If the pilot will prove to be successful, namely the deposition rate in OA repositories will reach approximately 60%, the OpenAIRE infrastructure will be continued in future EC programmes (FP8), richer content (research data) and greater volumes of publications and statistics will be managed, and other sub-programmes and disciplines, beyond the seven areas of the pilot, will take advantage of the infrastructure (e.g., the Future Emerging Technologies programme in bio-medical

sciences).

With respect to research (primary) data deposits, supplementing the text-publication, the challenge is somewhat more complicated. The nature of such materials is so diverse in different subjects, and their manipulation requires such different tools and facilities, that it will become indispensable to provide subject-specific functionality to be able to cope with the variety of requirements they put to the system, while tailoring to the specific needs of each discipline.

Finally, as the OpenAIRE infrastructure becomes more popular and deposition practices mature, the additional scenario of remote deposition, where researchers use OpenAIRE to deposit publications whose payload is automatically ingested in all authors repositories of reference, will be engineered and relative technical solutions will be devised.

Acknowledgments

This work is funded by FP7 Combination of Collaborative projects and Coordination and Support Actions - Integrated Infrastructures Initiative project (I3) proposal, with grant agreement: 246686 (call: FP7-INFRASTRUCTURES-2009-1). This work would have not been possible without the support, ideas and experience of partners and people from the OpenAIRE Consortium.

References

- ¹ Open Access Pilot in FP7, ftp://ftp.cordis.europa.eu/pub/fp7/docs/open-access-pilot_en.pdf, (Feb, 2010)
- ² UNESCO Open Access Handbook: Opportunities and Challenges, Published by the European Commission and the German Commission for UNESCO, 2008, http://dSPACE.org/images/publications/unesco%20handbook_open_access_english.pdf (2010)
- ³ a cooperation between DEFF, DFG, SURF Foundation and JISC set up in 2005, <http://www.knowledge-exchange.info/>
- ⁴ (2008). Recommendations from the EUA Working Group on Open Access. (http://www.eua.be/fileadmin/user_upload/files/newsletter/EUA_WG_open_access.pdf)
- ⁵ Open Access Pilot in FP7, ftp://ftp.cordis.europa.eu/pub/fp7/docs/open-access-pilot_en.pdf, (Feb, 2010)
- ⁶ The pilot covers approximately 20% of the Commission's FP7 research programme budget, and will run until the end of FP7 (2013).
- ⁷ ERC Scientific Council Statement on Open Access. <http://erc.europa.eu/pdf/open-access.pdf>, (Feb 2010)
- ⁸ D-NET Project: <http://www.d-net.research-infrastructures.eu>. Istituto di Scienza e Tecnologie dell'Informazione, Centro Nazionale delle Ricerche (Pisa, Italy), ICM Research Centre (Warsaw), University of Bielefeld Library (Bielefeld, Germany), and Department of Informatics, National and Kapodistrian University of Athens (Greece)
- ⁹ DRIVER Project: Digital Repository Infrastructure Vision for European Research, <http://www.driver-community.eu>
- ¹⁰ CDS Invenio <http://cdsware.cern.ch/invenio/index.html> (Feb 2010)
- ¹¹ Inventory study into the present type and level of OAI compliant Digital Repository activities in the EU, K. van Eijndhoven, M. van der Graaf <http://www.driver-support.eu/documents/DRIVER%20Inventory%20study%202007.pdf>
- ¹² <http://coar-repositories.org/>
- ¹³ SHERPA RoMEO: publisher copyright policies & self-archiving. <http://www.sherpa.ac.uk/romeo/> (Feb 2010)
- ¹⁴ <http://www.peerproject.eu/>
- ¹⁵ Repository for "orphan" researchers, i.e. researchers that do not have an institutional or thematic repository of reference.
- ¹⁶ Europeana Connecting Cultural Heritage Project, <http://www.europeana.eu> (2010)
- ¹⁷ Enabling Services in Knowledge Infrastructures: The DRIVER Experience Leonardo Candela, Donatella Castelli, Paolo Manghi and Pasquale Pagano, Proceedings of the Third Italian Research Conference on Digital Library Systems (IRCDL), Padua, Italy, 2007
- ¹⁸ D-NET Project: <http://www.d-net.research-infrastructures.eu>. Istituto di Scienza e Tecnologie dell'Informazione, Centro Nazionale delle Ricerche (Pisa, Italy), ICM Research Centre (Warsaw), University of Bielefeld Library (Bielefeld, Germany), and Department of Informatics, National and Kapodistrian University of Athens (Greece)
- ¹⁹ Ali Arsanjani, Service-oriented modeling and architecture, IBM developerWorks, <http://www-128.ibm.com/developerworks/webservices/library/ws-soa-design1>
- ²⁰ W3C Web Services Activity web site, <http://www.w3.org/2002/ws>
- ²¹ European Film Gateway: <http://www.europeanfilmgateway.eu> (Feb 21010)
- ²² The PIRUS2 Project, (JISC), <http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/tiki-index.php> (Feb 2010)
- ²³ ANSI/NISO Z39.93 - The Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol, http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=2de0e3e04f3a7e32d45db8ee87574c3c8206ddcb (Feb 2010)
- ²⁴ <http://repositories.webometrics.info/>
- ²⁵ <http://www.ebi.ac.uk/>
- ²⁶ <http://www.embl.de/>
- ²⁷ <http://www.mad.zmaw.de/>
- ²⁸ <http://www.mpimet.mpg.de/en/home.html>
- ²⁹ www.biodiversityinternational.org
- ³⁰ <http://www.cit-ec.de/>
- ³¹ <http://www.dans.knaw.nl/en/>



I search



I find



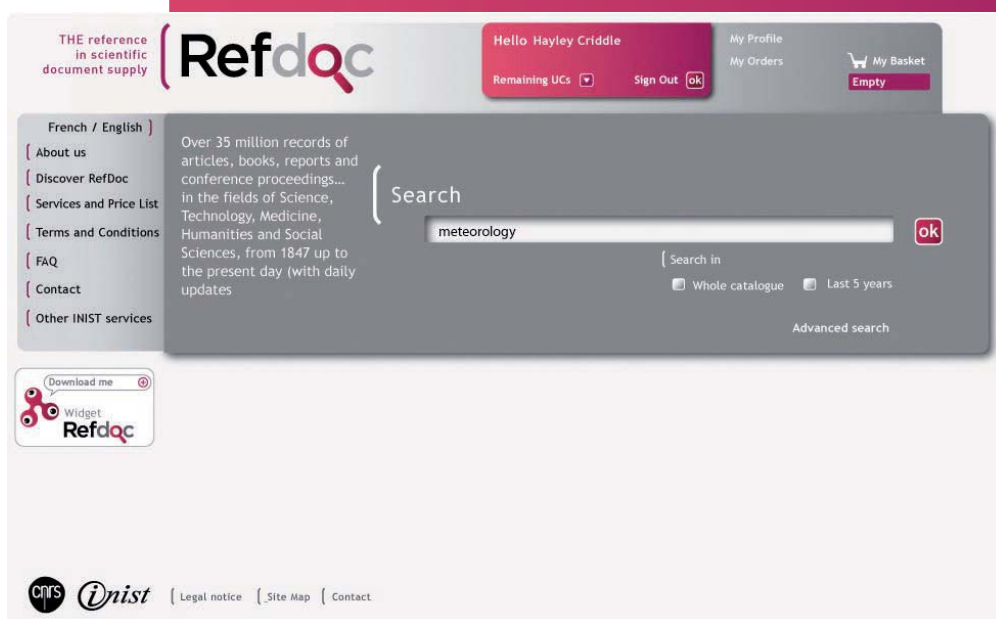
I order

January 2010 – all scientific
and technical information
available in just 3 CLICKS

Over 35 million records
of articles, books,
reports and conference
proceedings from 1847
up to the present

Refdoc.fr

DISCOVER ALL THE FEATURES AND FUNCTIONS AVAILABLE AT www.refdoc.fr



OpenSIGLE - Crossroads for Libraries, Research and Educational Institutions in the field of Grey Literature*

Dominic J. Farace and Jerry Frantzen (*Netherlands*),
Christiane Stock, Nathalie Henrot, and Joachim Schöpfel (*France*)

Abstract

This article is based on a paper¹ presented at the Tenth International Conference on Grey Literature (GL10) in which GreyNet's collections of conference preprints were made accessible via the OpenSIGLE Repository. OpenSIGLE offers a unique distribution channel for European grey literature with roots dating back a quarter century. In the first part of the article, the experience of INIST as service provider and GreyNet as data provider will be discussed including recent developments. Later in the article, the draft of a project proposal called for in the final session of that conference will be elaborated. The proposal seeks to explore the capacity required for the OpenSIGLE Repository to develop in multilateral and international cooperation in support of European research infrastructures committed to the open access of grey literature collections and resources. Emphasis is placed on the involvement of libraries, research centers, and institutions of higher education, as well as, requirements for a grey literature network service to sustain further development, exploitation, and promotion of the OpenSIGLE Repository.

From SIGLE to OpenSIGLE: A Progress Report

SIGLE (System for Information on Grey Literature in Europe) was a unique multidisciplinary database dedicated to grey literature. Up to 15 European partners participated in SIGLE, mostly national libraries or libraries aligned to well-known research institutes. Their principal goals were the centralized collection of scientific and technical reports, theses and other grey material and to facilitate access to these documents through an engagement for document delivery or loan. Created in 1980 and produced from 1984 onwards by EAGLE (European Association for Grey Literature Exploitation), the database was last available through STN International and on CD-ROM via Silverplatter/Ovid until it became dormant in 2005. INIST then decided to make the data publicly available on an open access platform. Details of the migration from SIGLE to OpenSIGLE have been presented at the GL8 Conference² held in December 2006 (Schöpfel 2007). And in December 2007, the OpenSIGLE website³ went live.

This article further discusses three related issues dealing with OpenSIGLE: (1) usage statistics covering two years of access to the repository, (2) a bilateral cooperative agreement with GreyNet, the Grey Literature Network Service, and (3) a project proposal exploring the capacity required for the OpenSIGLE Repository to develop in multilateral and international cooperation.

OpenSIGLE Traffic Report

Usage information for a database is at all times interesting for the producer of the information. In this case an additional incentive was the fact that OpenSIGLE records, which migrated from the SIGLE database, had not been updated since 2005. Would then the move to an open access environment be at all "useful" for the grey literature community?

The usage analysis is based on data obtained through phpMyVisites, an open source software for website statistics that works with a javascript image call. Only completely uploaded pages are counted and robots are excluded. The following data provide only a part of the information that can be obtained through phpMyVisites. Other statistics based on server logs might however provide even higher figures.

The first figure shows that the number of visits as well as the number of page views has increased steadily since the opening of the website in 2007. A first peak was reached in July 2008 following a press campaign in the middle of the French holidays. The result is both surprising and rewarding since visits usually decrease during summer months.

The usage of OpenSIGLE continues to increase and has more than tripled between August 2008 and August 2009 in terms of page views and number of visits, where the average duration of a visit is 100 seconds. Visits where only a single page is viewed represent a stable 50% average of the traffic to the site. These users accessed the database after

* First published in the GL11 Conference Proceedings, March 2010

searching via Google or Google Scholar. While in other cases, users may carry out extensive searches and view hundreds of web pages.

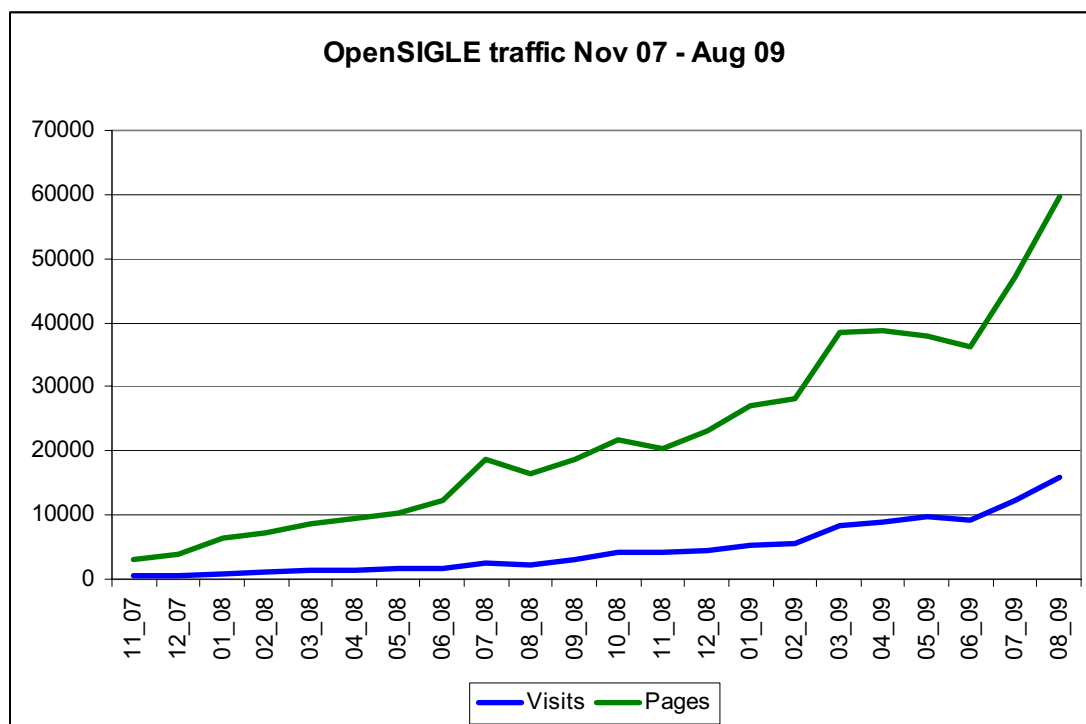


Figure 1: OpenSIGLE traffic report – number of visits and pages viewed

Geographic Origin of Visitors

The software used allows us to monitor the origin of visitors for the top ten countries each month. The sum of 22 months worth of data shows the United Kingdom in the lead, closely followed by the United States. A second group is formed by Germany, France, Italy and Spain. Countries in the long tail may not appear on a given monthly top ten listing. It is obvious that OpenSIGLE users come not only from the former EAGLE countries, but also from the United States, Canada, and since recently China and Australia. Clearly an indication that European grey literature presents an interest worldwide.

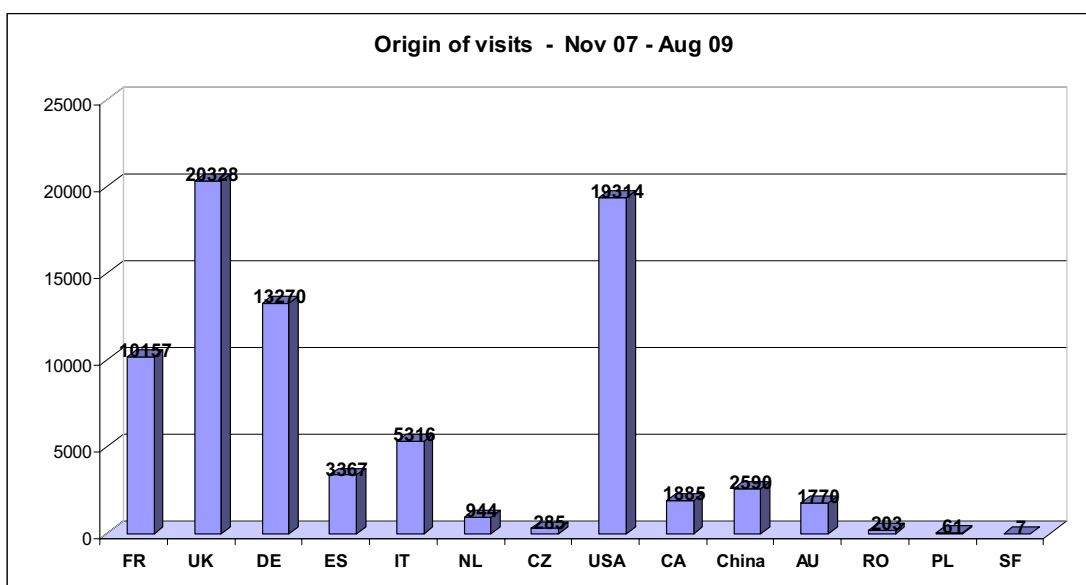


Figure 2: Origin of visitors to OpenSIGLE

Usage and Feedback

Compared to other INIST websites and e-resources, statistics show that 16% to 19% of the users come from North America. OpenSIGLE is in third place among users from this continent preceded by the English version of INIST's institutional website⁴ and IndicaSciences⁵ - an INIST product dedicated to research evaluation and indicators. INIST websites geared to a French speaking audience receive an average of 7% of the visits from North America.

The analysis of web links as well as feedback through incoming messages reveal that OpenSIGLE is often used in the biomedical and public health sectors. However, at present, statistics do not allow us to go into further detail regarding scientific domains.

During the course of 2008, several requests were received from former users of the STN or Ovid versions of the SIGLE database dealing with complex search strategies. Such questions required another look into the limits of the Jakarta Lucene search engine implemented within DSpace, especially with regard to the length of the search query. It was discovered that Lucene allows for more possibilities than mentioned in the help provided by DSpace. Besides inquiries involving search strategies, users were also interested in the download and export features of OpenSIGLE.

One critical view of OpenSIGLE found on a blog⁶, mentions the absence of links to the full text of documents. Of course this is understandable given the fact that it was one of the very reasons why the SIGLE database was discontinued.

Promotional Activities

Before the official announcement of the launch of OpenSIGLE, the project was presented at a DSpace meeting focused on the exchange of experiences among its users (Grésillaud and Stock, October 2007)⁷. Shortly afterwards, and as a result of that meeting, visitors from Spain and Italy were observed on the OpenSIGLE website. In December 2007, INIST also focused attention on OpenSIGLE during the GL9 Information Walk-Thru at the Ninth International Conference on Grey Literature in Antwerp, Belgium⁸.

In May 2008, a short presentation for the French public was given at I-expo (IT conference and exhibit) in Paris. And in July, INIST sent a press release to national and international lists and agencies *i.e.* Information World Review and Research Information. This no doubt resulted in the above mentioned peak of visits in the middle of summer. A brief message about OpenSIGLE was placed simultaneously on the French and international homepages of INIST. Since "news items" are normally less frequent during summer months, the message remained for a longer period of time on these WebPages.

Today OpenSIGLE is indexed by Google and Google Scholar and included in the bookmarks of national libraries and research institutes. Following the creation of the WorldWideScience Alliance and website⁹ in June 2008, INIST (a partner in this Alliance) proposed to integrate OpenSIGLE into the WorldWideScience portal. This was realized in September 2008. And, in the web statistics that following month WWS.org appeared as forth partner site for visitors accessing OpenSIGLE through a website with GreyNet.org¹⁰ following closely behind. Overall, these different promotional activities have had a positive impact on the use and branding of OpenSIGLE.

GreyNet, On the Background and Forefront of OpenSIGLE

In this section, the relationship between GreyNet and the former EAGLE Association including its SIGLE database will be addressed. This will then be followed by a conscious positioning of GreyNet in the newfound OpenSIGLE Repository with INIST as its Service Provider.

In 1992, EAGLE agreed to act as main sponsor for the launch of the International Conference Series on Grey Literature first held in the Amsterdam RAI in December 1993. GreyNet was at that time a newly established network service - driven on two fronts: (1) to promote the field of grey literature and the work of organizations involved in this branch of information the world over, and (2) to stimulate research on grey literature and make the results available both in print and digital (electronic) formats. EAGLE participated as sponsor and/or program committee member in the first five Conferences in the GL-Series.

In early 2005 GreyNet was invited as an observer to the final EAGLE Board meeting at FIZ Karlsruhe upon which the EAGLE Association formally voted to be dissolved. It was at that same meeting that the initial draft of an OpenSIGLE proposal¹¹ was presented by Dr. Joachim Schöpfel, last in line of EAGLE Presidents.

In the two ensuing years (2005-2007), INIST worked unilaterally on OpenSIGLE, which could then be described as a caretaker repository. In the autumn of 2007, once OpenSIGLE had become operational, GreyNet met with colleagues at INIST to hammer out an agreement that on the one hand would make GreyNet OAI-compliant and on the other

hand would expand INIST's role in OpenSIGLE from solely a caretaker to an external service provider. To this end, GreyNet's conference based collections would provide an example of OpenSIGLE's potential for other data providers in the grey literature community.

GreyNet's Collections in OpenSIGLE

In December 2008, five years of research issuing from the GL Conference Series had been uploaded in the OpenSIGLE Repository. The bilateral contact between INIST as service provider and GreyNet as data provider was successful in customizing a metadata record for the enriched publication of conference preprints and the subsequent migration of GreyNet's collections to an open access environment. The bilateral agreement likewise holds for future conferences in the GL-Series, continuing with GL10 records onward. Retrospective input of the initial four conferences in the GL-Series (1993-1999) would of course make GreyNet's collections comprehensive in OpenSIGLE. To this end, in January 2009, GreyNet purchased from Emerald Group Publishing – former MCB University Press – the rights to allow the full-text papers from the earlier four conferences in the GL-Series to be made available in the OpenSIGLE Repository. This step was not only applauded by the open access community¹², but it also suggests other possibilities to retrieve content controlled by commercial publishers¹³. GreyNet proceeded with the production of metadata records, while INIST took on the work of scanning and creating image files for the retrospective records. In October 2009, half of the retrospective input had been achieved.

GreyNet's Potential for OpenSIGLE

The initial reaction from the grey literature community to GreyNet's alliance with OpenSIGLE has been positive; however, due to the brief timeframe in which GreyNet's collections are actually available in the OpenSIGLE Repository, it is too early to provide substantial user statistics. While GreyNet has been receiving monthly reports from INIST generated via OpenSIGLE, GreyNet is looking for other ways to compile use and user statistics via its own channels. In this way, there would be separate data issuing from INIST as service provider and GreyNet as data provider that would allow for comparisons and provide grounds for decision making in the future.

In September 2008, an OpenSIGLE webpage was added to the GreyNet website with hyperlinks to its conference collections already in the repository; and in January 2009 that webpage became a main page on GreyNet's website. Not only did the number of visits to the webpage double in the first half of 2009, but it now also allows for the addition of sub-pages used for promotional and instructive purposes.

The Grey Literature Network Service feels that it has even more to offer OpenSIGLE than its conference collections. Going back to 1992, when GreyNet was first launched, one of its primary goals was to promote the field of grey literature and the work of organizations involved in this branch of information. What EAGLE was to SIGLE, GreyNet could be to OpenSIGLE and more. GreyNet operates internationally and maintains a full-time established network service specializing in grey literature with information products and resources both in print and electronic formats. GreyNet has for the past seven years (2003-2009) often together with colleagues from INIST carried out research projects involving citation analysis, surveys, interviews, as well as standard review of the literature. Over the past years (1992-2009), GreyNet has developed channels for promotional outreach as well as a modest publishing arm. More recently, GreyNet has set up a program of training and instruction in the field of grey literature, which could also be linked to OpenSIGLE. These and other such initiatives would no doubt serve and support future developments in the OpenSIGLE Repository.

OpenSIGLE Project Proposal, A Feasibility Study

What began unilaterally with the vision and determination of INIST and what has recently been expanded in bilateral cooperation with GreyNet has yet even greater potential for the international grey literature community. GreyNet together with INIST are committed to drafting a project proposal. This proposal will explore the capacity required for the OpenSIGLE Repository to further develop in multilateral and international cooperation in support of European research infrastructures committed to open access of their grey literature collections and resources. One in which special emphasis is geared to libraries, research centers, and institutions of higher education.

Project Lead-Time

Both INIST and GreyNet have put forth a number considerations and recommendations based on their recent experience with the OpenSIGLE Repository. An inventory of issues and recommendations were collated and will be used in the development of work packages

in the implementation phase of the project. Some of the issues include: closing time gaps in record entries, linking to full-text documents as well as plus links to datasets and software, integrating OpenSIGLE in other networks and portals, streamlining the SIGLE Classification scheme, etc.

Project Consortium

Based on the main objective of the proposed project and in relation to the issues that would have to be dealt with in order to achieve this objective, project partners and external advisors need to be identified and brought together in a consortium for the duration of the project. To achieve optimal results, the number of stakeholders in the project would be limited. In the diagram below, the content as well as management base of the project is visualized. However, the names of the prospective organizations, who would be carrying out the projects' roles and tasks are masked here until final confirmation.

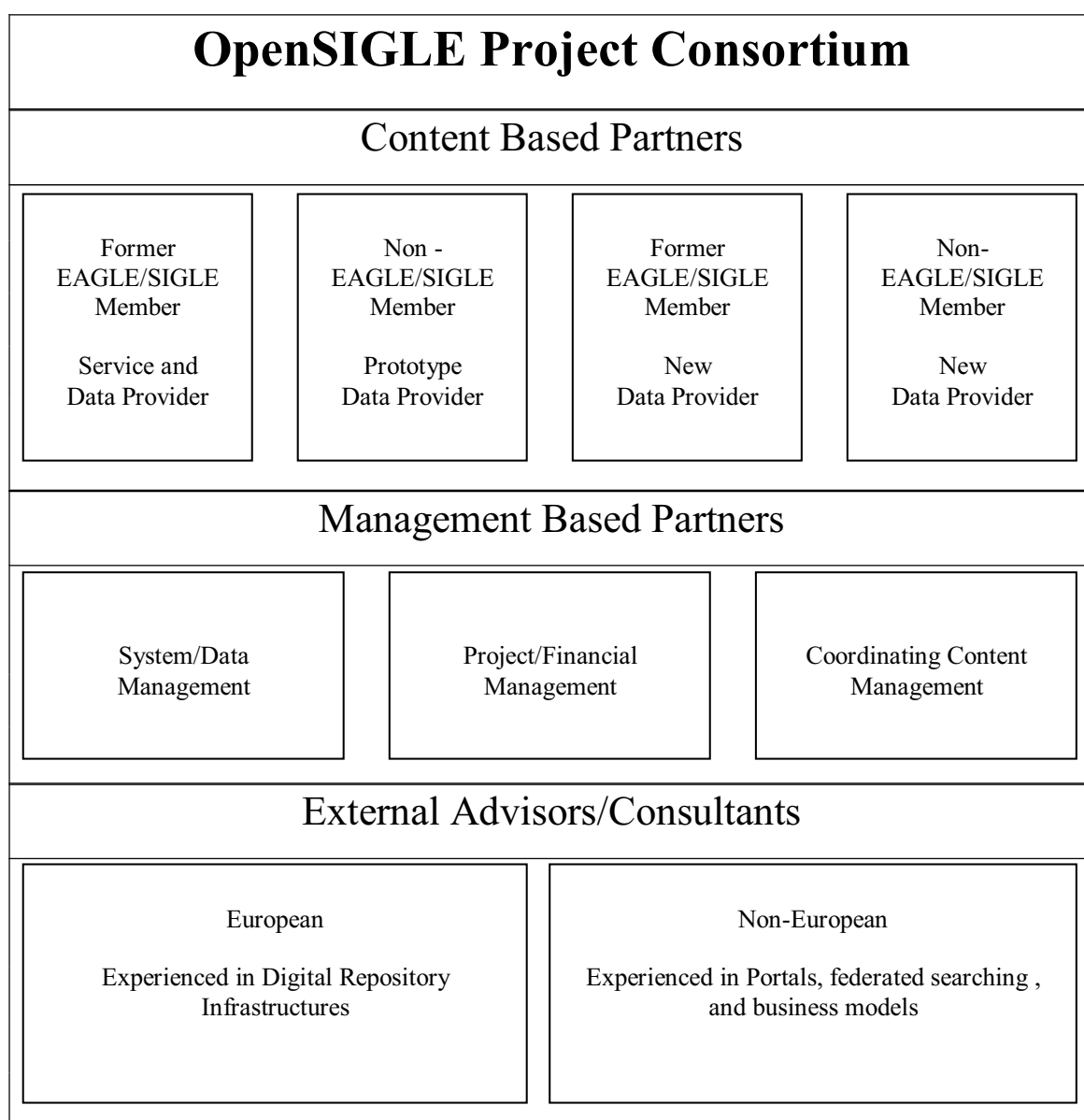


Figure 3: OpenSIGLE Project Consortium

Expected Results and Impact of the Project

This project has its roots in a European framework of cooperation among longstanding infrastructures including national libraries, research centers, and networked services. The outcome of this project would support and strengthen policy development for infrastructures in the field of grey literature, where open access to their collections and other knowledge based resources stand central. The OpenSIGLE Repository with its technical know-how would be sustained by a coordinating infrastructure in the advancement of European cross-disciplinary research well beyond its geographical borders. A draft of this project proposal will be presented during a Panel Session at the Eleventh International Conference on Grey Literature¹⁴ held in the Library of Congress in Washington D.C on 14-15 December 2009. The panel members will take the opportunity to discuss the project proposal in order to illicit feedback from the international grey literature community, raise public awareness to the OpenSIGLE Repository, and solicit leads for further project funding.

Appendix: Timeline SIGLE – OpenSIGLE

2009-...	2006-...	2003-...	1999-...	1985-...
2009 GreyNet's Retrospective Collections 1997-1999 in OpenSIGLE; GL11 -Eleventh International Conference on Grey Literature in Washington D.C. USA with INIST as Co-Sponsor 2008 OpenSIGLE Repository included in WWS.org, World-Wide Science Portal; GreyNet's Collections 2003-2007 accessible via OpenSIGLE Repository; GL10 – Tenth International Conference on Grey Literature in Amsterdam with INIST as Co-Sponsor 2007 OpenSIGLE operational as caretaker repository; Bilateral agreement INIST (Service provider) – GreyNet (Data Provider); GL9 – Ninth International Conference on Grey Literature in Antwerp, Belgium with INIST and the EU as Co-Sponsors	2006 SIGLE migration to OpenSIGLE with INIST as Service Provider; GL8 – Eighth International Conference on Grey Literature in New Orleans with INIST as Co-Sponsor 2005 EAGLE Association dissolved and SIGLE Database dormant; INIST proposal for OpenSIGLE; GL7 – Seventh International Conference on Grey Literature in Nancy, France with INIST as Host and EU as Co-Sponsor 2004 GL6 – Sixth International Conference on Grey Literature in New York, NY with INIST as Co-Sponsor	2003 GreyNet relaunch; GL5 – Fifth International Conference on Grey Literature in Amsterdam with INIST as Co-Sponsor and EAGLE on Program Committee 2000 GreyNet discontinued	1999 GL'99 – Fourth International Conference on Grey Literature in Washington D.C. USA 1997 GL'97 – Third International Conference on Grey Literature in Luxembourg with EC as Host and EAGLE on Program Committee 1995 GL'95 – Second International Conference on Grey Literature in Washington D.C. USA; 1993 GL'93 – First International Conference on Grey Literature in Amsterdam, EAGLE Sponsor; Bi-Annual GL-Conference Series 1993-1999 with EAGLE on the Program Committee 1992 GreyNet founded	1985 EAGLE Association was founded as producer of the SIGLE Database 1980 SIGLE established as database and document delivery system; Later exploited via STN and Blaise Hosts as well as SilverPlatter CD-ROM 1978 York Seminar on Grey Literature hosted by British Library and the EC

References

- ¹ Farace, D.J., J. Frantzen, C. Stock, N. Henrot, and J. Schöpfel (2009), OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal. – In: *The Grey Journal : An International journal on Grey Literature*, vol. 5, no 1, Spring 2009. ISSN 1574-1796
- ² Schöpfel, J., C. Stock, and N. Henrot (2007), From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource Migration in the European Context. – In: *The Grey Journal : An International journal on Grey Literature*, vol. 3, no 1, Spring 2007. ISSN 1574-1796
- ³ OpenSIGLE - System for Information on Grey Literature in Europe, <http://opensigle.inist.fr/>
- ⁴ English version of INIST's institutional website, <http://international.inist.fr>
- ⁵ INIST product dedicated to research evaluation and indicators, <http://indicasciences.inist.fr>
- ⁶ Critical view of OpenSIGLE found on a blog, <http://healthinformaticist.wordpress.com/2008/08/28/does-opensigle-exist-for-its-own-sake/>
- ⁷ Grésillaud, S., and C. Stock (October, 2007), DSpace at INIST-CNRS: one platform, different usages and resulting specific needs/problems. Paper presented at DSpace User Group Meeting 2007, Food and Agriculture Organization of the United Nations, Rome, Italy. Available at <http://www.aepic.it/conf/viewabstract.php?id=208&cf=11>
- ⁸ Grey Foundations in Information Landscape (2007), Ninth International Conference on Grey Literature, 10-11 December 2007 in Antwerp, Belgium. – GL9 Conference Program and Abstracts. – ISBN 978-90-77484-09-8
- ⁹ WorldWideScience.org, the global science gateway, <http://worldwidescience.org/>
- ¹⁰ GreyNet, Grey Literature Network Service, <http://www.greynet.org/>
- ¹¹ Schöpfel, J. (2006), MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE. – In: GL7 Conference Proceedings, pp. 34-39. – ISBN 90-77484-06-X
- ¹² Posting by Peter Suber on January 29 (2009), <http://www.earlham.edu/~peters/fos/2009/01/greynet-buys-rights-to-deposit-papers.html>
- ¹³ Posting by Heather Morrison on August 6 (2009), <http://www.connotea.org/comments/uri/92b11113ecf827be19a369f21e81161b>
- ¹⁴ GL11 Program and Conference Bureau, <http://www.textrelease.com/>

Twelfth International Conference on Grey Literature

National Technical Library, Prague, Czech Republic, 6-7 December 2010



Transparency in Grey Literature
Grey Tech Approaches to High Tech Issues

Call-for-Papers

GUIDELINES FOR THE SUBMISSION OF ABSTRACTS

Participants who seek to present a paper at GL12 are invited to submit an English abstract between 300-400 words. The abstract should deal with the problem/goal, the research method/procedure, an indication of costs related to the project, as well as the anticipated results of the research. The abstract should likewise include the title of the proposed paper, name(s) of the author(s), and full address information. Abstracts are the only tangible source that allows the Program Committee to guarantee content and balance in the conference program. Every effort should be made to reflect the content of your work in the abstract submitted. Abstracts not in compliance with the guidelines will be returned to the author for revision.

TITLE OF PAPER, AUTHOR AND ADDRESS INFORMATION

Author Name(s)	Telephone
Organization(s)	Fax
Address/P.O. Box	Email
Postal Code - City - Country	URL

INDICATE THE THEMATIC THREAD MOST SUITED TO YOUR ABSTRACT

- ☐ **Peer Review** - Quality Control - Enriched Publication
- ☐ **Redefining Grey Literature** - Proof of Concept - Research Process
- ☐ **New Stakeholders** - Subject and Community based approaches - Geographic Perspectives
- ☐ **Copyright** - Licensing - Authors Rights - Open Access - Open Source
- ☐ **Standardization** - Interoperability - Metadata - Repositories
- ☐ **Economics of Grey** - Management and Sustained Resources - Value Added Services

DUE DATE AND FORMAT OF SUBMISSION

Abstracts must be emailed by **April 30, 2010** in MS Word. The author will receive verification upon receipt. Shortly after the Program Committee meets in May 2010, those who submitted abstracts will receive notification of their place on the GL12 Conference Program. This notice will be accompanied by further guidelines for the submission of the full-text papers and PowerPoint presentations. Authors will then have 30 days to complete their Conference Registration. For further details, see the GL12 Conference Registration Form.

Jonathan Cain is a MSLIS Candidate at the School of Information Sciences, Pratt Institute. He holds a BS from the College of Charleston and a MA in from New York University. He serves as a Reference Associate for the Sciences at New York University's Bobst Library. His current research interest focuses on government information accessibility, Web 2.0 technologies for intelligence gathering and diplomacy, and models of information access in societies with low internet access. Email: jccain30@pratt.edu

Dominic J. Farace founded GreyNet, Grey Literature Network Service, in 1993 and has since been responsible for the international GL-Conference Series. In this capacity, he serves as Program and Conference Director as well as managing editor of the conference proceedings. Since 2004, he is a Guest Lecturer on Grey Literature in the Masters Program at the University of Amsterdam; Instructor of Grey Literature via UNO Distance Education, and Editor of TGJ. Email: dominic.farace@textrelease.com

Tina Gheen is the Library Director at the National Science Foundation (NSF) in Arlington, Virginia, where she manages library operations, projects, and staff. She is particularly interested in making the grey literature in the NSF Library collection machine readable and searchable so that it may be leveraged for future scientific work. She received her MLS from the University of North Texas and her BS in Biology from Southwest Texas State University. She is currently a representative of CENDI, an interagency working group of senior scientific and technical information managers from 13 U.S. federal agencies; FLICC, the Federal Library and Information Center Committee; the FEDLINK Advisory Council; and the Naval Research Laboratory Alliance. Email: tgheen@nsf.gov

Wolfram Horstmann is Chief Information Officer (CIO) for Scholarly Information at Bielefeld University. He is also responsible projects such as OpenAIRE, in which Bielefeld University is working on data management aspects in the D-NET infrastructure as well as studies on disciplinary requirements and usage statistics. He was formerly involved in the DRIVER-II project and worked as Scientific-Technical Manager of the first DRIVER project at the State and University Library of Göttingen and was head of the department for publishing services at the academic library centre "hbz" in Cologne. He has a scientific background in Computational Neuroscience and has received his PhD for work in the Theory of Science on the role of simulations in brain sciences. Email: wolfram.horstmann@uni-bielefeld.de

Paolo Manghi, Research Fellow at the NMIS lab of the Istituto di Scienza e Tecnologie dell'Informazione (ISTI), Consiglio Nazionale delle Ricerche, Pisa, Italy. His interests are in the fields of Data Models for Digital Libraries, Types for Compound Objects in Digital Repositories, Digital Library Systems and Services, Service-Oriented Infrastructures for Digital Libraries, Database Systems, Type systems for XML languages, Query languages for XML data, XML P2P database systems. Member of the core expert group of Europeana and of the DL.org workgroup on content interoperability. He is involved in the DRIVER/DRIVER II, OpenAIRE and EFG EC projects and in the Microsoft project R2D2. Email: Paolo.Manghi@ISTI.CNR.IT

Natalia Manola is a Senior Software Engineer holding a B.Sc. in Physics from the University of Athens, Greece, and an M.Sc. in Electrical and Computer Engineering from the University of Wisconsin at Madison, USA. Her professional experience consists of several years of employment as a Software Engineer, Software Architect, Information Technology Administrator, and Information Technology Project Manager by companies in various Information Technology sectors in the US and Greece. Recently she was responsible for design and development of User Interface components of the D-NET Software Toolkit, in the projects DRIVER, DRIVER-II and today OpenAIRE. The systems she

has designed and implemented include specialized ETL tools, biotechnology and genetic applications, embedded financial monitoring systems, heterogeneous data integration systems, and service-oriented architectures. She has also participated and technically managed several R&D projects funded by the European Union or by the national government.

Maliaca Oxnam has been a science-engineering librarian at the University of Arizona Libraries (UAL) since 2000. Her leadership roles include leading several strategic projects for UAL, including the Strategic Long Range Planning for the Libraries. Maliaca has also served as a co-PI on the development of the Geotechnical Rock & Water Resources Digital Library (GROW) and as a key participant on Einstein's Protégé's": A Heuristic approach to Bridging the Gap between Faculty Expectations and Student Preparedness funded respectively by the National Science Foundation and the Department of Education. Maliaca has served as vice-chair/chair-elect of the ALA ACRL Science & Technology Section and serves as chair of the Technical Report Archive & Image Library (TRAIL) Taskforce on a joint project for the Greater Western Library Alliance and Center for Research Libraries aimed at enabling and increasing access to US scientific GL. maliaca@u.arizona.edu

Leonid P. Pavlov graduated from Moscow Physical-Engineering Institute, Dipl. Eng. in computer systems. He is a Candidate of Sciences in informatics; and since 1976 is employed with the Scientific and Technical Information Centre of Russia (VNTIC) as Deputy Director. Main works in information systems, scientific and technical information, and grey literature. Email : pavlov@vntic.org.ru

Dale Peters is working since 2009 at the Academic Computing ICT division of the University of KwaZulu-Natal. Formerly, she was Scientific Technical Manager of the DRIVER project, based at the University Library Göttingen, and directly responsible for the collaboration between DRIVER and related projects, including PEER and EUROPEANA. In the past Dale Peters provided leadership to DISA: Digital Innovation South Africa, in the collaborative development of digital library services across multiple remote institutions through a central digital library infrastructure. She is a consultant for UNESCO, IFLA, and the Library of Congress, participating in regional and international initiatives aimed at the development of digital repositories in support of emerging paradigms of electronic scholarship. Email: petersd@ukzn.ac.za

Joachim Schöpfel obtained his Ph.D. in psychology from the Hamburg University in 1992. From 1991 to 2008, he worked at the French Institute for Scientific and Technical Information (INIST-CNRS) in different positions in database production and library management, at last as head of the e-publishing and document supply department. During the same time, he was lecturer at the University of Nancy. At present, he is senior lecturer in information and communication sciences at the Charles de Gaulle University of Lille 3. He published on GL, document delivery, digital libraries, scientific publishing, usage statistics and professional development. Email: joachim.schoepfel@univ-lille3.fr

Christiane Stock is the Head of the Monographs and Grey Literature service at INIST, in charge of the repositories LARA (reports), mémSIC (master's theses in information sciences) and OpenSIGLE. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). She is member of the AFNOR expert group who prepared the recommended metadata scheme for French electronic theses (TEF). Email: christiane.stock@inist.fr

Notes for Contributors

Non-Exclusive Rights Agreement

- I/We (the Author/s) hereby provide TextRelease (the Publisher) non-exclusive rights in print, digital, and electronic formats of the manuscript. In so doing,
- I/We allow TextRelease to act on my/our behalf to publish and distribute said work in whole or part provided all republications bear notice of its initial publication.
- I/We hereby state that this manuscript, including any tables, diagrams, or photographs does not infringe existing copyright agreements; and, thus indemnifies TextRelease against any such breach.
- I/We confer these rights without monetary compensation and with the understanding that TextRelease acts on behalf of the author/s.

Submission Requirements

Manuscripts should not exceed 15 double-spaced typed pages. The size of the page can be either A-4 or 8½x11 inches. Allow 4cm or 1½ inch from the top of each page. Provide the title, author(s) and affiliation(s) followed by your abstract, suggested keywords, and a brief biographical note.

A printout or PDF of the full text of your manuscript should be forwarded to the office of TextRelease. A corresponding MS Word file should either accompany the printed copy or be sent as an attachment by email. Both text and graphics are required in black and white.

REFERENCE GUIDELINE

General

- All manuscripts should contain references
- Standardization should be maintained among the references provided
- The more complete and accurate a reference, the more guarantee of an article's content and subsequent review.

Specific

- Endnotes are preferred and should be numbered
- Hyperlinks need the accompanying name of resource and date; a simple URL is not acceptable
- If the citation is to a corporate author, the acronym takes precedence
- If the document type is known, it should be stated at the close of a citation.
- If a citation is revised and refers to an edited and/or abridged work, the original source should also be mentioned.

Examples:

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

Jeffery, K.G., A. Asserson, J. Revheim (2000), Grey Literature and the knowledge Society. - In: Proceedings CRIS-2000, Helsinki, Finland. ftp://ftp.cordis.lu/pub/cris2000/docs/jeffery_fulltext.pdf

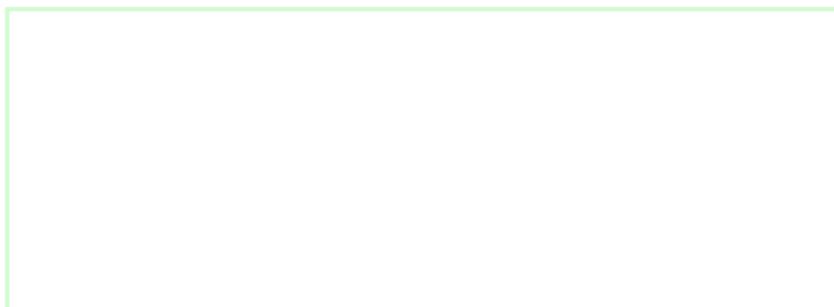
DCMI, Dublin Core Metadata Initiative Home Page http://purl.oclc.org/metadata/dublin_core/

Review Process

The Journal Editor first reviews each manuscript submitted. If the content is suited for publication and the submission requirements and guidelines complete, then the manuscript is sent to one or more Associate Editors for further review and comment. If the manuscript was previously published and there is no copyright infringement, then the Journal Editor could direct the manuscript straight away to the Technical Editor.

Journal Publication and Article Deposit

Once the journal article has completed the review process, it is scheduled for publication in The Grey Journal. If the Author indicated on the signed Rights Agreement that a preprint of the article be made available in GreyNet's Archive, then browsing and document delivery are immediately provided. Otherwise, this functionality is only available after the article's formal publication in the journal.



An International Journal on Grey Literature
'GOVERNMENT ALLIANCE TO GREY LITERATURE'
 SPRING 2010 – TGJ VOLUME 6, NUMBER 1

A Multi-Institutional Approach to Technical Report Literature: Development of the Technical Report Archive & Image Library (TRAIL)	5
Maliaca Oxnam (<i>United States</i>)	
Digitizing Grey Literature from the <i>Antarctic Bibliography</i> Collection	9
Tina Gheen and Sue Olmsted (<i>United States</i>)	
Web 2.0 Utilization in E-Diplomacy and the Proliferation of Government Grey Literature	15
Jonathan Omowalé Cain (<i>United States</i>)	
The Grey System for Monitoring Self-Funded Research	25
Aleksandr V. Starovoitov, Yuri M. Bogdanov, Aleksandr M. Bastrykin, and Leonid P. Pavlov (<i>Russia</i>)	
An Infrastructure for Managing EC Funded Research Output: The OpenAIRE Project	31
Paolo Manghi (<i>Italy</i>), Natalia Manola (<i>Greece</i>), Wolfram Horstmann (<i>Germany</i>), and Dale Peters (<i>South Africa</i>)	
OpenSIGLE - Crossroads for Libraries, Research and Educational Institutions in the field of Grey Literature	41
Dominic J. Farace and Jerry Frantzen (<i>Netherlands</i>), Christiane Stock, Nathalie Henrot, and Joachim Schöpfel (<i>France</i>)	

Colophon.....	2
Editor's Note.....	4
On the News Front	
GL12 Call for Papers - Twelfth International Conference on Grey Literature.....	49
Transparency in Grey Literature, Grey Tech Approaches to High Tech Issues	
Advertisements	
NTK – Czech National Technical Library.....	8
IIa – Information International Associates, Inc.....	24
J-STAGE – Japan Science and Technology Agency.....	30
Refdoc.fr – Institute for Scientific and Technical Information.....	40
About the Authors.....	50
Notes for Contributors.....	51