Grey Foundations in Information Landscape

NINTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

House of the Province, Antwerp, Belgium, 10 -11 December 2007

Conference Host

Department of Economy, Science and Innovation, Flemish Government
Belgium

GL9

Conference Proceedings

ISSN 1386-2316

Conference Sponsors

EBSCO
Information Services
United States

GESIS
German Social Science Infrastructure Services
Germany

INIST-CNRS
Institute for Scientific & Technical Information
France

NYAM
New York Academy of Medicine
United States

OPOCE
Office for Official Publications of the European Communities
Luxembourg

Program Committee

EWI	Department of Economy, Science and Innovation, Flemish Government	Belgium
CNR	National Research Council	Italy
euroCRIS	Current Research Information Systems	Norway
GESIS	German Social Science Infrastructure Services	Germany
GreyNet	Grey Literature Network Service	Netherlands
INIST-CNRS	Institute for Scientific and Technical Information	France
OSTI-DOE	Office of Scientific and Technical Information	USA
UCI	University of California, Irvine	USA

Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
www.textrelease.com · conference@textrelease.com
T/F +31 (0) 20-331.2420

CIP

GL9 Conference Proceedings
Ninth International Conference on Grey Literature : Grey Foundations in Information Landscape, 10-11 December 2007, House of the Province, Antwerp, Belgium / ed. by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2008. – 153 p. ; 30 cm. – Author Index. – (GL-conference series, ISSN 1385-2316 ; No. 9)

INIST-CNRS and NYAM are corporate authors and associate members of GreyNet. These conference proceedings contain the full text of over twenty papers presented during the two days of Plenary, Panel, and Poster Sessions. The papers appear in the same order as the conference program. Included are a List of Participating Organizations, Sponsored Advertisements, and an Author index.

ISBN 978-90-77484-10-4

Foreword

"Grey Foundations in Information Landscape"

The Ninth International Conference on Grey Literature sought to map the infrastructure in which grey literature is embedded. This concerted drill in the field of information also stands to further a framework for shared understanding. GL9 surveyed the information landscape by exploiting newfound and existing grey resources, by utilizing tools developed in part or whole for processing grey literature, and by clearly demonstrating uses and applications for research and policy driven settings. The title of the conference "Grey Foundations in Information Landscape" encompassed five main themes:

- Tools for Publishing, Archiving, and Accessing Grey Literature
- Use and Impact of Grey Literature in Scholarly Communication
- Grey Literature in Central and Eastern Europe
- New Discoveries in Grey Literature for Research Communities
- Education and Grey Literature

The conference venue provided information professionals with a variety of platforms for presenting and communicating their research results. Plenary and Panel sessions were held in the Antwerp Provincial Council Hall. And, in the meeting area adjacent to the lobby, an 'Information Walk-Thru' accommodated Product and Service Reviews as well as Poster presentations.

On behalf of the Conference Host and Sponsors, the Program Committee, and Chairpersons, I would like to thank the authors and co-authors for their content contributions to these proceedings. Likewise, I welcome those reading these conference proceedings to voice their comments and/or recommendations either directly to the authors or to GreyNet, Grey Literature Network Service.

Finally, I would like to bring to your attention the recent Call-for-Papers for **GL^{10}**, the Tenth International Conference on Grey Literature, which will be held in Science Park Amsterdam on 8-9 December 2008.

Dr. Dominic J. Farace
Grey Literature Network Service

Amsterdam,
February 2008

GL9 Conference Sponsors

EWI, Belgium
Department of Economy, Science and Innovation;
Flemish Government

EBSCO, USA
EBSCO Information Services

GESIS, Germany
German Social Science Infrastructure Services

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

NYAM, USA
The New York Academy of Medicine

OPOCE, Luxembourg
Official Publications of the European Communities

GL9 Program Committee Members and Organizations

Dr. Geert Van Grootel ***(Chair)***
Department of Economy, Science and Innovation, Belgium

Dr. Joachim Schöpfel
Institut de l'Information Scientifique et Technique, France

Agnieszka Wenninger
GESIS Service Agency Eastern Europe;
Social Science Information Centre, Germany

Daniela Luzi
National Research Council, Italy

Dr. Dominic J. Farace
Grey Literature Network Service, Netherlands

Anne Asserson
euroCRIS, Current Research Information Systems, Norway

Deborah Cutler
Office of Scientific & Technical Information, Dept. of Energy, USA

Julia Gelfand
University of California, Irvine, USA

Contents

OPENING SESSION
GREY FOUNDATIONS IN INFORMATION LANDSCAPE

Chair: Dr. Geert Van Grootel, Dept. of Economy, Science & Innovation; Flemish Government, Belgium

SESSION ONE
TOOLS FOR PUBLISHING, ARCHIVING, AND ACCESSING GREY LITERATURE

Chair: Julia Gelfand, University of California, Irvine Libraries, United States

SESSION TWO
USE AND IMPACT OF GREY LITERATURE IN SCHOLARLY COMMUNICATION

Chair: Anne Asserson, University of Bergen, Norway

SESSION THREE
GREY LITERATURE IN CENTRAL AND EASTERN EUROPE

Chair: Agnieszka Wenninger, GESIS Service Agency Eastern Europe, Germany

SESSION FOUR
NEW DISCOVERIES IN GREY LITERATURE FOR RESEARCH COMMUNITIES

Chair: Dr. Joachim Schöpfel, INIST-CNRS, France

PANEL SESSION
EDUCATION AND GREY LITERATURE

Chair: Daniela Luzi, CNR-IRPPS, Italy

POSTERS PRESENTATIONS,
PRODUCT AND SERVICE REVIEWS

Why wasn't the iPod invented in Flanders?

Why doesn't Flanders boast many innovative products? How can we best solve this lack of innovation?

The department of Economy, Science and Innovation (EWI) tries to find answers to this problem and to dozens of other questions like these.

Each day our staff works at forging a link between these three fields with the goal of making Flanders a durable, creative and enterprising region.

EWI

Combining Economy, Science and Innovation for a better society

www.ewi-vlaanderen.be

Greyscape

Keith G Jeffery
Science and Technology Facilities Council
Rutherford Appleton Laboratory (United Kingdom)

Anne Asserson
University of Bergen (Norway)

Abstract
This paper addresses the place of grey not only in the landscape of information, but also of knowledge, wealth creation and improvement of the quality of life. It builds on work by the authors published in previous GL conferences.

The very notion of grey is linked conceptually with dull and dismal. It refers to a concept between recognised states. It is also linked with age and distinction, even the magical (Gandalf the Grey). The hypothesis is that grey objects exhibit all these properties, and they can be used to advantage.

Typically Grey literature consists of technical research documentation (although many different kinds of material have been classified as grey). This – by its nature – tends to be dull. The material is usually not peer-reviewed as is white literature. Yet it is not usually the unstructured, unauthorised output from a single source; it may well be commercial in confidence and certainly is likely to contain IP (intellectual property) of value to the organisation. It is between states i.e. between 'an idea' and 'white'. Some grey literature has age and distinction, although this usually is neither recognised nor appreciated widely until it becomes white – analogously to the final step in the human ageing process. Finally some grey literature is indeed magical in providing an organisation with a wealth of knowledge for problem solving, strategy inspiration or know-how.

This paper attempts to demonstrate that grey is the very foundation of the knowledge base both for internal use of an individual organisation (where it may well remain grey) and of world knowledge (where it is transformed to white). The information landscape (which when applied becomes the knowledge landscape) has exposed uplands (white) and hidden valleys (grey).

The dynamics of the landscape concern the way in which an idea, concept or knowledge is generated and transformed: from grey to grey (internal discussion within an organisation with improvement), from grey to white (publication, public relations for an organisation, improved evaluation scores for an organisation), from grey to product or service (wealth creation or improvement in the quality of life within an organisation), from white to product or service (wealth creation or improvement in the quality of life by knowledge or technology transfer).

The prerequisites for these dynamics to work are (1) excellent metadata (to improve discovery and control usage), (2) an institutional document repository of grey, (3) an institutional CRIS for the contextual research information, (4) linkage between the document repository and the CRIS of an institution and thence (in a controlled manner with formal descriptive and restrictive metadata) to other institutions, (5) an e-research repository of research datasets and software, (6) linkage between the e-research repository and the CRIS of an institution and thence (in a controlled manner with formal descriptive and restrictive metadata) to other institutions, (7) an institutional policy to mandate deposition of the material with appropriate metadata.

From foundation to dynamics, grey is the key ingredient for the knowledge society.

BACKGROUND

Previous papers on Grey literature (or more widely – and used hereafter to mean - grey objects such as hypermedia) by the authors (in the GL Conference Series) have described:

1. the need for formal metadata to allow machine understanding and therefore scalable operations (Jeffery 1999);
2. the enhancement of repositories of grey (and other) e-publications by linking with CRIS (Current Research Information Systems) (Jeffery and Asserson 2004);
3. the use of the research process to collect metadata incrementally reducing the threshold barrier for end-users and improving quality in an ambient GRIDs environment (Jeffery and Asserson 2005);
4. an architectural model for scaleable, highly distributed, workflowed repositories of grey literature based on hyperactive 'intelligent' documents (Jeffery and Asserson 2006).

This paper takes a 'from 10,000 metres altitude' view of the grey information landscape (hence 'Greyscape') based on the hypothesis that grey literature is the foundation for the knowledge economy. The paper develops recommendations for realizing the potential of the hypothesis.

THE HYPOTHESIS

The hypothesis is that grey literature can provide the foundational material for knowledge transfer leading to wealth creation and improvement of the quality of life, i.e. the 'knowledge economy'. Furthermore, we propose that the dynamics of 'grey' must be captured as the grey object (e.g. technical report, PhD thesis, brochure) transforms during its life from grey to grey or grey to white. We also propose that the grey object must be linked via semantic relations not only to other grey objects and white objects but also to information on persons, organizations, funding, projects, facilities, equipment, products, patents, events… in fact all the contextual information required to understand the relevance of the grey object of interest. The current state of the art does not support this vision yet but is moving in this direction. We propose the Greyscape Architecture to solve the problems and provide a unifying platform for the world of grey to reach its potential.

THE NOTION

The concept of 'grey' leads to associations in various dictionaries and thesauri. These include:

- dull and dismal: as in a grey day implying clouds and rain or to describe the character of a person;
- obscured: implying only partly seen as in 'greyed out';
- between states: as in a grey area implying some uncertainty or fuzziness such as not clearly legal but neither strictly illegal;
- age and distinction: as implied by grey hair indicating ageing and wisdom;
- magical: as in Gandalf the Grey or associated with the colour of the cloaks of some religious orders;

In fact these characterizations apply well to grey literature. Commonly embracing material such as technical reports, catalogues, brochures, ephemera (but also doctoral and master's theses), learning materials, photographs, video and audio clips the epithet 'dull' is probably apposite. Commonly grey literature is obscured – hidden away, poorly catalogued, not advertised and difficult to find. It usually is considered less important than white, published literature and receives less attention. Certainly grey literature may be between states; the contents of a technical report (grey) or a doctoral thesis (grey) may well be transformed into a published research paper (white). Similarly there may be a transition from grey to grey (hopefully with added information / improvement). Grey literature may acquire the characteristics of age and distinction; early technical reports may be highly acclaimed once they are recognised and there are instances of old doctoral theses being cited and used once the author has started to publish white research papers on the topic. Interestingly, are old photographs of Einstein with equations written on the blackboard considered as grey literature? Finally some grey literature is indeed magical; a technical report may well - for a company – lead to a product making large sales and profits.

This paper addresses the IT systems architecture required to maximise the value of grey literature in wealth creation and improvement of the quality of life.

STATE OF THE ART

Despite some interesting counter-examples – mainly as research ideas or prototypes - in general the world of grey literature is characterized by: limited digitization (i.e. much on paper); limited fill of repositories (i.e. even if the grey object is digital it is stored in a user location without appropriate metadata); various repositories of material with different characteristics; different standards for metadata to describe or catalog the material; different query languages and capabilities; differing facilities to present the results. Furthermore, there is a lack of integration with repositories of white literature, research datasets and software and a lack of integration with CRIS which provide contextual information. Finally the different grey literature repositories commonly do not interoperate (or do not interoperate effectively). The bottom line is that grey literature is not sufficiently indexed by metadata and even when it is so indexed, the metadata are insufficient to provide sufficient context, provenance and other linkages.

This is a somewhat grey (in the sense of dismal) landscape. However, it can be improved rapidly and effectively by bringing together the excellent – but isolated - work that has been done into an integrated architecture allowing for heterogeneity dictated by local requirements but managing the heterogeneity for interoperability benefit.

REQUIREMENT

Easy to deposit

The first requirement – without which no grey literature repositories exist – is that it must be easy to deposit grey objects with associated metadata. This is best achieved by reducing the threshold barrier for the end-user; appropriate techniques include workflow and incremental metadata provision, automated metadata extraction and re-use and knowledge-assisted metadata input so reducing the amount of actual input required from the end-user.

Easy to retrieve

The major purpose of a grey literature repository is to provide grey objects upon request; the request usually being for one or more 'full text' objects (it may of course be hypermedia) of which the metadata satisfies the criteria of the search. Exceptionally a service may be provided for full-text search of the object itself for a word or phrase; however this is expensive and time-consuming and is much less easy to do when the grey object contains hypermedia including images, video and audio. It is clear that the relevance (accuracy, precision) of the response to the request and the recall (completeness) of the response both depend critically on the quality of the metadata. Furthermore, the request may have a scope not only of the local repository but global; in this case excellent metadata is required for effective interoperability achieved by transforming the request into the syntax (character set, language, grammar or structure) and semantics (meaning, especially meaning in context) of each target grey source

Easy to transition

Grey objects usually record a stage or step in a research process. They are related to earlier grey material and white material. They are related to future grey and white material. The relationships or linkages are critically important to understanding the relevance of a grey object to the request (or more specifically to the intent of the request). Moreover tracking the evolution of research thought through a time-series of grey and white objects is instructive. It may also be important legally in claims of prior publication and other rights.

Easy to track provenance

Related to the above, provenance information may improve the ability of the end-user to assess the quality and relevance of the grey object of interest. Provenance includes not only predecessor objects but also the contextual information mentioned above. It is important to have correct recording of versions of a grey object for the same reasons. Finally, for provenance information to be provided it is essential that facilities for preservation and curation are provided.

Easy to relate

From the foregoing it is clear that relationships (hyperlinks) are of the greatest importance for effective and efficient provision of grey resources. The relationships concern not only other objects in the grey repository (which should be open access and institutional) but also in the CRIS (which provides the research context) and also in research repositories of datasets and software. The relationships require:

- to have a consistent structured syntax for automation (hence precision, accuracy and scaleability) of their management and usage and for interoperation;
- to be annotated with consistent semantic information so that they may be interpreted automatically by computer to achieve scalability;

An example might be as follows:

<person> 'Einstein, Albert' <datetimestart> '19160000:0000' <datetimeend> '19160000:0000' <role> 'lecturer' <publication> ' Einstein, Albert (1916). "Photograph of lecture 1916"

<person> 'Einstein, Albert' <datetimestart> '19160000:0000' <datetimeend> '19160000:0000' <role> 'author' <publication> ' Einstein, Albert (1916). "The Foundation of the General Theory of Relativity" Annalen der Physik.'

<publication> ' Einstein, Albert (1916). "The Foundation of the General Theory of Relativity" Annalen der Physik.' <datetimestart> '19160000:0000' <datetimeend> '19160000:0000' <role> 'presented' <publication> ' Einstein, Albert (1916). "Photograph of lecture 1916"

<person> 'Einstein, Albert' <datetimestart> '19050630:0000' <datetimeend> '19050926:0000' <role> 'author' <publication> 'Einstein, Albert (1905) "Zur Elektrodynamik bewegter Körper" Annalen der Physik 17(1905), 891-921. Received June 30, published September 26, 1905.'

<publication> ' Einstein, Albert (1916). "The Foundation of the General Theory of Relativity" (PDF). Annalen der Physik.' <datetimestart> '19160000:0000' <datetimeend> '19160000:0000' <role> 'theory improvement' <publication> 'Zur Elektrodynamik bewegter Körper, Annalen der Physik 17(1905), 891-921. Received June 30, published September 26, 1905.'

<person> 'Einstein, Albert' <datetimestart> '19030000:0000' <datetimeend> '19110000:0000' <role> 'employee' <organisationalunit> 'Swiss Patent Office Bern'

<person> 'Einstein, Albert' <datetimestart> '19030000:0000' <datetimeend> '19050000:0000' <role> 'postgraduate' <organisationalunit> 'University of Zurich'

This example could be extended with other grey and white publications, Einstein's undergraduate studies at ETH Zurich, his famous 'thought experiments' (as projects), subsequent books and papers by other authors citing his work and other entities.

This example demonstrates clearly several aspects required of metadata:

- the syntax must be formal and precise;
- the semantics must be present, formal and precise (usually using a domain ontology to assure a controlled structure of terms and inter-relationships of meaning);
- the relationships form a fully-connected graph; hierarchic representations are insufficient. This excludes XML (XML) in its 'native' form; if XML represents relationships as relations (rather than using its native linkage mechanism XLINK) then the full syntax can be reconstructed;
- the relationships require an annotation richer than the triples of RDF (RDF) so once again RDF is insufficient (although RDF overcomes the primary limitations of XML);

From this we conclude that the information should be represented using a relational database system although it may be presented (with a reduction in dimensions of the information space) on web pages at the user interface via XML. Furthermore, a logic programming environment is required to manage the semantics (derived from the domain ontology) in the expressions.

THE PROPOSED ARCHITECTURE

The proposed architecture has three major components:

- a CRIS to provide formal metadata (for scaleable and quality interoperation) and to provide contextual information;
- a repository of grey objects (and ideally white objects) and ideally associated (via the CRIS) with a repository of research datasets and software;
- a workflow system providing incremental metadata update through the research process to reduce the threshold effort to get material into the repository;

This proposal is no more than an integration of separate aspects proposed in previous papers by the authors and discussed under 'Background' above.

We reject at this stage (but will re-introduce later) the architecture based on hyperactive grey objects (Jeffery and Asserson 2006) because for many existing grey resources hosts the technology is just too advanced and the required e-infrastructure is commonly missing at institutions concerned with grey literature. However, technology based on integration of repositories using CRIS, and associated processes to fill the repositories using workflow, are reasonably well understood, accepted and implemented.
This leads to the architecture proposed for implementation at any one institution (Figure 1):

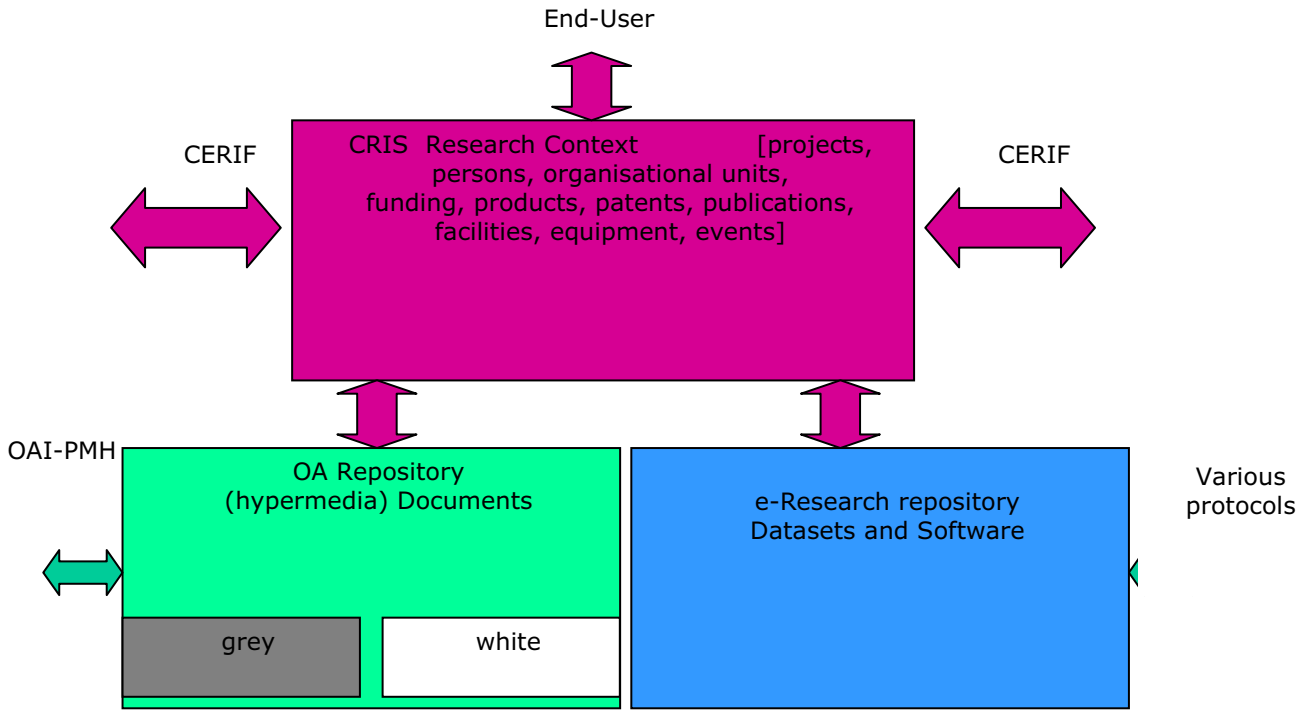


Figure 1: Architecture at One Institution

It should be noted that this architecture puts the CRIS 'centre stage'. This is because of its more formal syntax and semantics (CERIF). This formality is required for optimal management of the information in the repositories and the CRIS and especially in the interlinking, expressed as time-period-stamped, role-based relationships. This (CERIF) metadata also provides an optimal method for interlinking multiple institutions, each with the CRIS and repositories as indicated for one institution. Indeed, the proposal is that CERIF can provide a better (more formal, more complete, more descriptive) interoperation mechanism than OAI-PMH (for repositories of publications) or the many different protocols for interoperation of research repositories of datasets and software (Figure 2).

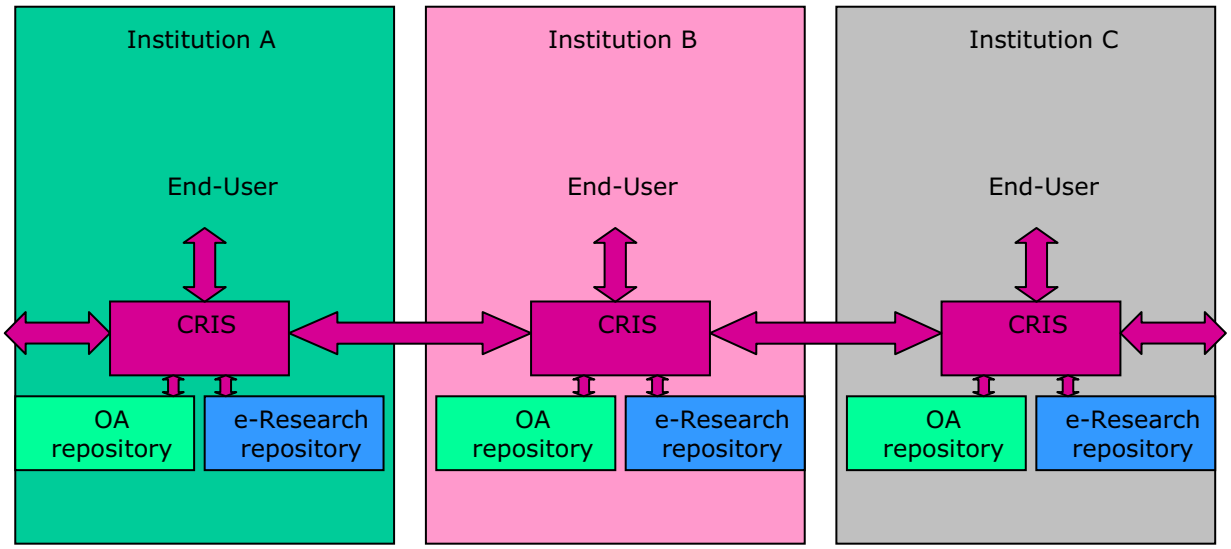


Figure 2: Architecture at multiple institutions

A STEPWISE APPROACH TO GREYSCAPE

We propose a set of steps or workpackages to achieve Greyscape.
(1) discussion and agreement on a standard for excellent (formal syntax and semantics) metadata (to improve discovery and control usage);
(2) an institutional document repository of grey set up at each participating organization;
(3) an institutional CRIS for the contextual research information set up at each participating organization;
(4) linkage between the document repository and the CRIS of an institution (using a unique identifier attached to metadata generated and stored in the CRIS and replicated to the repository) and thence (in a controlled manner with formal descriptive and restrictive metadata) to other institutions,
(5) an e-research repository of research datasets and software set up at each participating organization;
(6) linkage between the e-research repository and the CRIS of an institution and thence (in a controlled manner with formal descriptive and restrictive metadata) to other institutions as well as to the repository of grey objects,
(7) an institutional policy to mandate deposition of the material with appropriate metadata (to encourage filling of the repositories so providing more and better information to support research and innovation);
(8) Information management, analysis and prediction services for the CRIS and the two repositories to provide end-users with the information required for research, for finding collaborators, for finding innovative ideas or products, for assessing research and many other uses;
(9) Workflow processes to connect services and users to reduce effort for users with data re-use

When all of this version of Greyscape is in place, all interested organizations can move up together to utilizing – within the same basic e-infrastructure - hyperactive grey objects as proposed in (Jeffery and Asserson 2006); furthermore the architecture allows backward compatibility to the architectural version proposed here from any subsequent version based on hyperactive grey objects.

CONCLUSION

Now is the time for 'grey' to change its image from negative to positive and benefit wealth creation and improvement in the quality of life. A blueprint is proposed here to achieve Greyscape.

REFERENCES

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

(CERIF) www.eurocris.org/cerif

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

(DSpace) http://www.dspace.org/

(FRIDA) http://frida.uio.no

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

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

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

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

(Jeffery and Asserson 2004) K G Jeffery, A G S Asserson; Relating Intellectual Property Products to the Corporate Context; Proceedings Grey Literature 6 Conference, New York, December 2004; TextRelease; ISBN 90-77484-03-5

(Jeffery and Asserson 2005) K G Jeffery, A G S Asserson 'Grey in the R&D Process'; Proceedings Grey Literature 7 Conference, Nancy, December 2005; TextRelease; ISBN 90-77484-06-X ISSN 1386-2316

(Jeffery and Asserson 2006) Keith G Jeffery, Anne Asserson: 'Hyperactive Grey Objects' Proceedings Grey Literature 8 Conference, New Orleans, December 2006; TextRelease; ISBN 90-77484-08-6. ISSN 1386-2316 ; No. 8-06-X

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

(XML) http://www.w3.org/XML/

Accessing Grey Literature in an Integrated Environment of Scientific Research Information

Elly Dijk, Chris Baars, Arjan Hogenaar, and Marga van Meel
Royal Netherlands Academy of Arts and Sciences, KNAW (Netherlands)

Abstract
The DARE programme was initiated in the Netherlands by the SURF Foundation, the Dutch universities, the Royal Netherlands Academy of Arts and Sciences (KNAW), and the Netherlands Organisation for Scientific Research (NWO). The objective of the programme was to enable digital access to research results of all the Dutch scientific institutions. Since the DARE programme (2003 - 2006), all the institutes involved have had their own repositories.

The DAREnet portal website (www.darenet.nl/en) gives access to the content of all the repositories. As from 1 January 2007, KNAW Research Information has taken over responsibility for the DAREnet website, formerly maintained by the SURF Foundation. DAREnet currently contains three sections: DAREnet proper, Cream of Science, and Promise of Science. In November 2007 DAREnet provides access to some 138,000 digital publications at 19 institutions. Cream of Science showcases prominent research from the Netherlands. It lists over 200 top Dutch academics, providing worldwide access to more than 48,000 publications. About 60 percent of these can be accessed full-text (PDF). Promise of Science enables full-text search of over 16,000 e-theses. In all, some 30 percent of the objects can be listed as grey literature. These grey publications have always been hard to find but can now easily be accessed in DAREnet. They can also be found via Google and other search engines.

Another service run by of KNAW Research Information is NARCIS, the National Academic Research and Collaborations Information System. NARCIS (www.narcis.info) is a portal website where users can find research information (programmes, projects, researchers and their expertise, research institutes) from the Current Research Information System (CRIS), full-text publications, and news items from research institutes' websites. Later this year datasets will also be included. Because there is some overlap in content and functionality between the DAREnet and NARCIS systems, KNAW Research Information has decided to integrate DAREnet into NARCIS. This involves the creation of one back-end system and centralised content updates in one location. All the functionalities for current users of both portals will remain, including the RSS feed and advanced search options, and new ones will be added.

After the integration of DAREnet and NARCIS, the big challenge will be the interlinking of the three major research information types in NARCIS: programme and project descriptions, primary data, and research publications (articles, reports, books, patents, etc.).

This interlinking will be made possible by the development of an exchange format for the source systems. Unique digital author identifiers, a project in which each individual researcher in the Netherlands is given a unique number, are being used to create the links.

In the paper the development of connecting the information in NARCIS will be described in detail. NARCIS will be part of the Dutch Academic Information Domain. When completed this system will enable full-text (grey) literature and any underlying datasets to be searched in their context (research programme, institute, fields of expertise).

Finally, the European dimension, i.e. the DRIVER project - Digital Repository Infrastructure Vision for European Research (http://www.driver-repository.eu) will be described.

Introduction

The mission of the Royal Netherlands Academy of Arts and Sciences (KNAW) is to ensure the quality of scientific research in the Netherlands and to promote the open accessibility of scientific information. The main task of the department of Research Information is to be the national focal point for research information in the Netherlands.

For this purpose the Academy has been producing the Dutch Research Database (the NOD), which is the national Current Research Information System. This database contains information about current research, researchers and research institutes. The Academy is also responsible for the DAREnet website, which gives open access to Dutch full text publications, and NARCIS, a portal containing scientific information in the Netherlands. In NARCIS the information from the NOD and DAREnet is combined, and extended with scientific news items and datasets.

Within "Knowledge Exchange (KE)", a collaboration of DFG, JISC, DEFF and SURF, a new concept has been developed for connecting information from CRIS (Current Research Information Systems) and OAR (Open Access Repositories): the 'Academic Information Domain' (AID). In this paper the focus will be on the realisation of a Dutch Academic Information Domain.

THE PORTALS OF KNAW RESEARCH INFORMATION

The department of Research Information of the Academy is responsible for the Dutch Research database (NOD), DAREnet and NARCIS.

Dutch Research Database (NOD)

The Dutch Research Database (NOD)[i] is being produced by the department of Research Information of the KNAW. This database is the national Current Research Information System (CRIS) and offers an overview of the scientific landscape of the Netherlands. The NOD is a publicly available online database with information on scientific research (research programmes and research projects), researchers (with their working addresses and their expertise), and research institutes (with profiles). The database covers all scientific disciplines and gives access to university and non-university research information. The NOD is a relational database and the information is highly structured: it offers links between research, persons and institutes.

The NOD contains information on 36,000 researchers (names linked to research data) of which are 7,500 professors (with addresses and teachings), and 9,100 researchers with expertise (and addresses). One can find 37,000 descriptions of research programmes and projects, from which 20,000 are current. And there is information on 3,200 research institutes, of which 750 with a profile (mission).

Sources of the NOD are: (1) research information systems of universities and research institutes; (2) research information systems of funding organisations, like the Netherlands Organisation for Scientific Research (NWO); (3) overviews and annual reports of institutions; (4) information from the researcher; (5) websites, and (6) press and news reports.

The NOD is fully accessible by search engines like Google. This has lead to an average of 170,000 unique visitors, and 240,000 visits per month (Figure 1). About the half of all the users has accessed the NOD via a search engine.

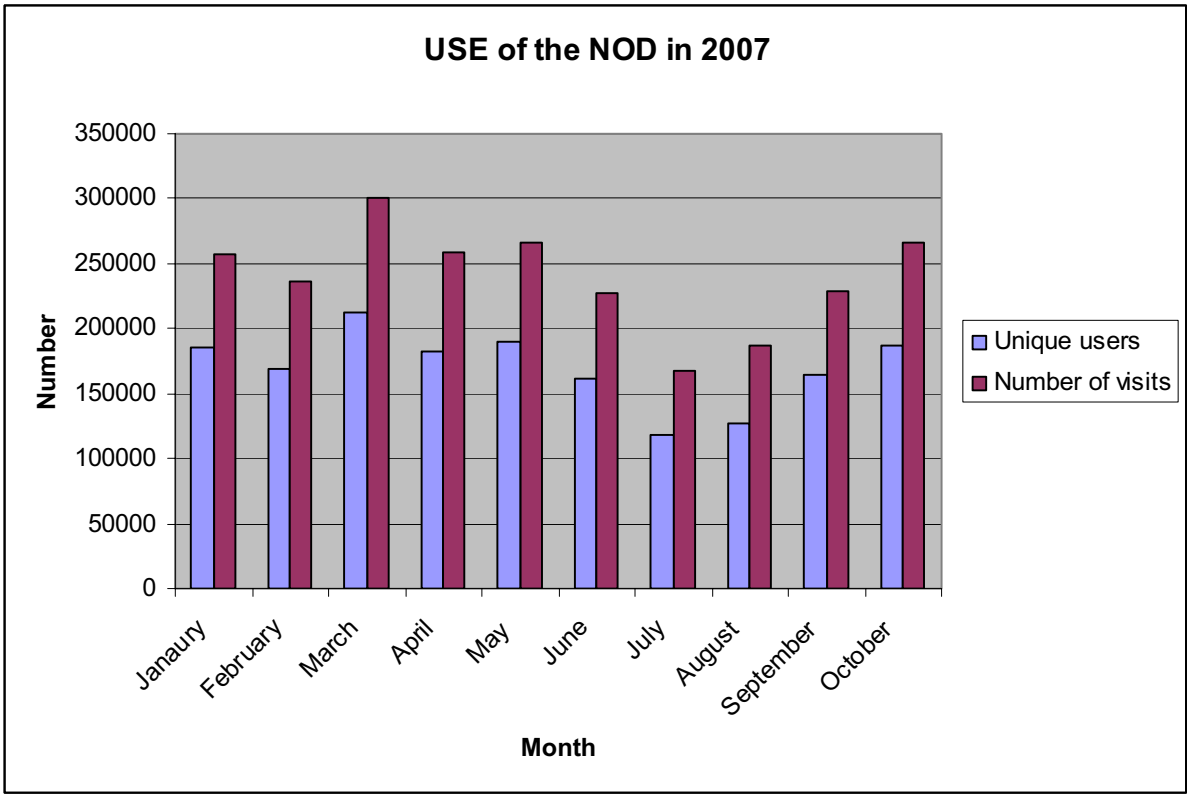


Figure 1. Use of the NOD in 2007

DAREnet

In the Netherlands the four-year DARE programme was launched in 2003. DARE is short for Digital Academic REpositories. The mission of the DARE programme was to get better access to results of publicly funded academic research in the Netherlands. It is a programme which allows authors to post their publications in an institutional academic repository. The DAREnet[ii] website gives access to these repositories. This is referred to as the so-called "green route of Open Access publishing".[iii] DAREnet gives access to Grey Literature as well. About 30% of the publications are grey literature.

DARE started with a budget of almost six million Euros. All the universities in the Netherlands are participants, as are the Academy, the Netherlands Organisation for Scientific Research as the most important funding organisation, and the National Library. This makes DARE a unique national programme. SURF - the ICT partnership organisation for higher education and research - was coordinating the project.

In the first year an infrastructure of institutional academic repositories was set up, based upon the Open Archives Initiative Protocol for Metadata Harvesting.

The first milestone of the project was reached in January 2004 when the demonstrator portal DAREnet was launched. At that time there were 17.000 records in DAREnet.

In 2004 and 2005 the focus was on populating the repositories. Within a year, almost 49,000 objects had been uploaded, not only the metadata but also the full text. In May 2005 the second milestone was reached: Cream of Science. Cream of Science is the showcase of Dutch prominent research. All DARE partners selected ten of their prominent academics whose complete publication lists were made visible and digitally available through DAREnet, with as much full text as possible. One of the prime aims of Cream of Science is to open up top quality content to the scientific community and society at large, and make it more easily and electronically accessible. Another aim is to demonstrate that scholars are willing to post their materials to a repository. In the end over 200 authors were willing to cooperate.

A different project was Promise of Science. The aim of Promise of Science was to set up a national doctoral e-thesis gateway as part of DAREnet and populate this gateway with 10,000 full text e-theses before the end of 2006. The total annual production of doctoral theses is around 2,500 items, which equals 5% of the formal scholarly output in the Netherlands.

In January 2007, when the DARE programme came to an end, the KNAW took over the responsibility of DAREnet.

At this moment (November 2008) DAREnet contains about 138,000 full text scientific publications and research output from all Dutch universities, some scientific institutes, KNAW and NWO. Cream of Science is containing 46,000 scientific publications written by 229 prominent Dutch scientists. About 60% of these can be accessed full text. Promise of Science gives access to over 16,000 doctoral e-theses from all Dutch universities.

In 2007 DAREnet has about 26,000 unique users and an average of 38,000 visits in a month (Figure 2).

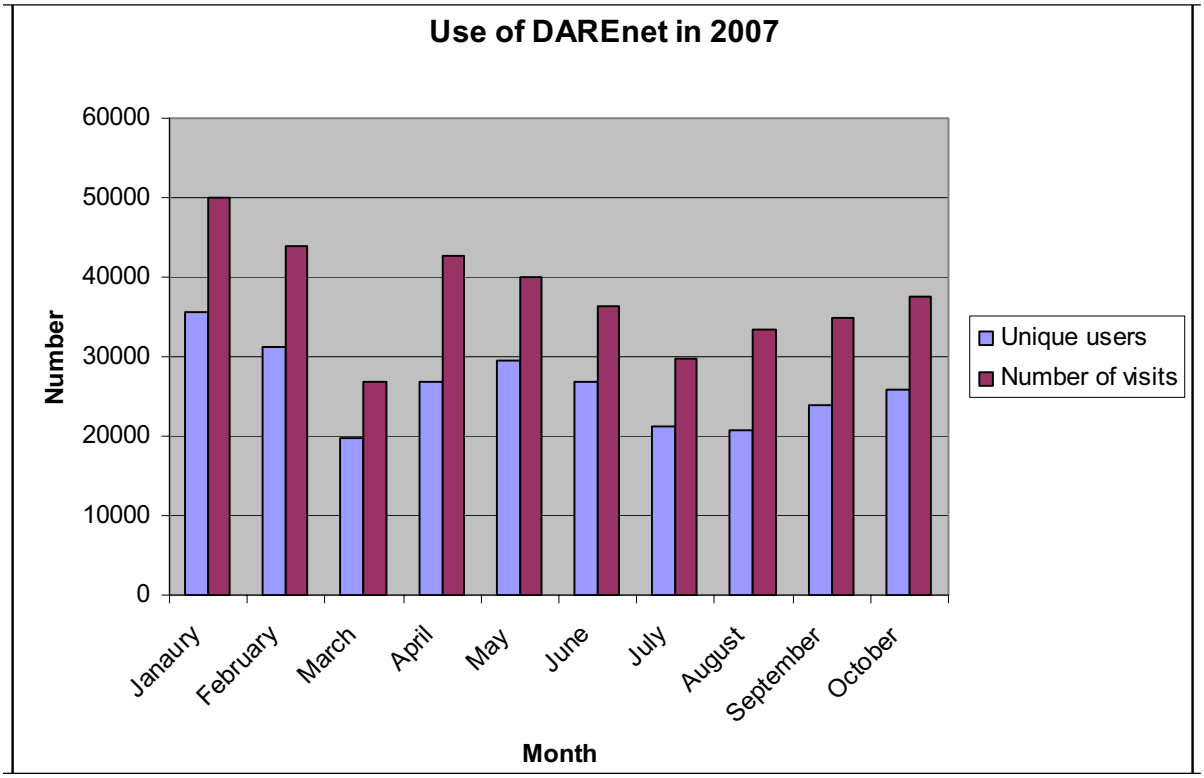


Figure 2. Use of DAREnet in 2007

(NB In March the statistical programme wasn't running properly)

NARCIS

The Academy has taken over the responsibility for DAREnet because of the great overlap with NARCIS[iv]. The NARCIS portal has been developed by the Research Information department of the Academy and harvests the academic repositories too - like DAREnet. This is the immediate cause to integrate NARCIS and DAREnet.

As part of the DARE programme, the Academy initiated the NARCIS project in close collaboration with NWO. NARCIS is short for National Academic Research and Collaborations Information System and has been set up in order to increase the visibility and accessibility of Dutch scientific research.

NARCIS combines information of DAREnet and the NOD (Dutch Research Database) together with the research datasets which are stored by DANS in their archive system EASY. DANS - Data Archiving and Networked Services - is an institute of KNAW and NWO, and is the national organisation responsible for storing and providing permanent access to research data from the humanities and social sciences. DANS has developed its own set of metadata, describing datasets. As from December 2007 the metadata of the research data of DANS are also being harvested in NARCIS.

Integration of DAREnet into NARCIS

In 2007 DAREnet was integrated into NARCIS. First of all the back-end system was integrated. This is the system that ensures that all content is brought together.

By using a so called OAI-PMH bridge the NARCIS harvester is able to harvest the metadata of all the CRIS records. The XML format used in the harvest consists of only those fields that are used for search functionality of the portal. The records that are actually shown are directly from the CRIS, which is an oracle database. So the OAI-PMH protocol serves only as a vehicle for the information stored in the index (Figure 3).

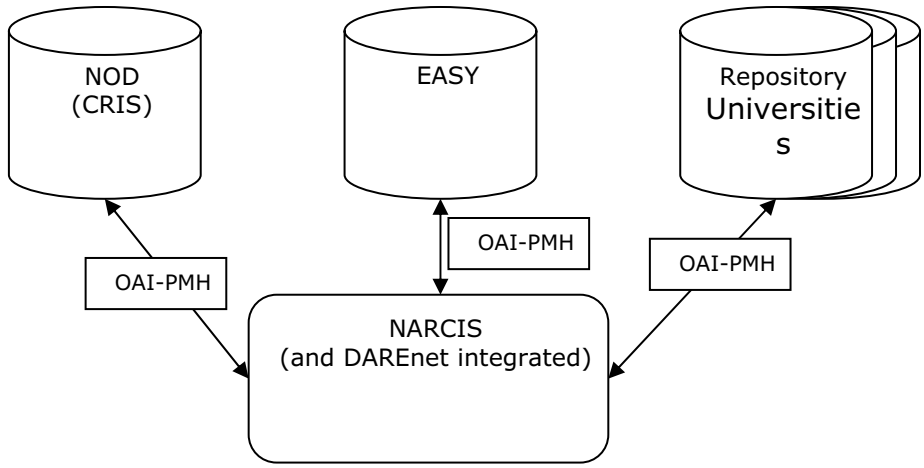


Figure 3. NARCIS harvest the national CRIS, EASY and the academic repositories

After the integration of the backend the department of Research Information started to integrate the interfaces (Figure 4). The advanced search option and the special access features to Cream of Science and Promise of Science in DAREnet have also become a part of NARCIS.

The new NARCIS home page now displays scientific news items from various sources (university websites, scientific news magazines) with the option of accessing the full articles. This news content is refreshed every hour. The home page also lists all universities, KNAW and NWO, again with the option of drilling down to underlying units, current research and scholars. All records have been assigned fixed URLs which may be used for reference purposes. In this new NARCIS version records can be found using search engines like Google. By means of an RSS feed NARCIS automatically informs its users about new records in their areas of interest. In 2007 there have been several releases of the portal to give the users the opportunity to utilize new functionalities. In January 2008 when the new portal will have a new look and feel, the integration of NARCIS and DAREnet will be ready.

Figure 4: the homepage of NARCIS on 31 October 2007: www.narcis.info

The use of NARCIS

Since September 2007 the records in NARCIS can be found by search engines like Google. It takes some time for the robots to visit the website, but in November, for the first time, there are much more visitors than in the months before. In January 2008 the website will be launched officially, so the expectation is that the use will increase afterwards. Until now NARCIS is only known by a limited group.

In DAREnet and the Dutch Research Database the use is much higher. On average DAREnet is visited by 26,000 unique users a month and the Dutch Research Database has 170,000 unique visitors per month. So we expect the number of users of NARCIS, where the information from the Dutch Research Databases and DAREnet is integrated, to grow rapidly.

NARCIS and Grey literature

Like DAREnet, NARCIS also gives access to grey literature, from universities and some other research institutes. Scientific grey literature comprises newsletters, reports, working papers, theses, government documents, bulletins, fact sheets, conference proceedings and other publications distributed free, available by subscription, or for sale.[v] In NARCIS about 30% of all the publications is grey literature for example theses, research reports, technical reports, conference papers, et cetera.

The problem with grey literature was always that it was hard to find. Now, with the rise of repositories, it is possible for scientific institutions to ingest the full text publications in their open access repositories, what will make it easy to find them. DAREnet and NARCIS are having ten of thousands of users each month. It is not possible to distinguish the users of grey literature, not all the universities can give these figures, but of course a part of the use is downloading the grey literature publications.

THE DUTCH ACADEMIC INFORMATION DOMAIN

Within "Knowledge Exchange (KE)"[vi], a collaboration of DFG, JISC, DEFF and SURF, a new concept has been developed for connecting information from CRIS (Current Research Information Systems) and OAR (Open Access Repositories): the 'Academic Information Domain' (AID). As can been seen in Figure 5, the Academic Information Domain can be distinguished from the Personnel Information Domain or the Financial Information Domain[vii].

CRIS are information systems containing an extensive set of metadata covering various aspects of research information. They are in use at universities, research institutes, and/or governmental bodies and were initially developed for administrative purposes. OAR contain for example full text publications – including Grey Literature – and research data (in the so-called e-Research repository[viii]). Many of the repositories are managed at a single-institution level.

NARCIS can be seen as the Dutch Academic Information Domain with information from the national CRIS, the NOD, and the full text publications from the academic repositories, and the research data in the field of the humanities and social sciences of DANS.

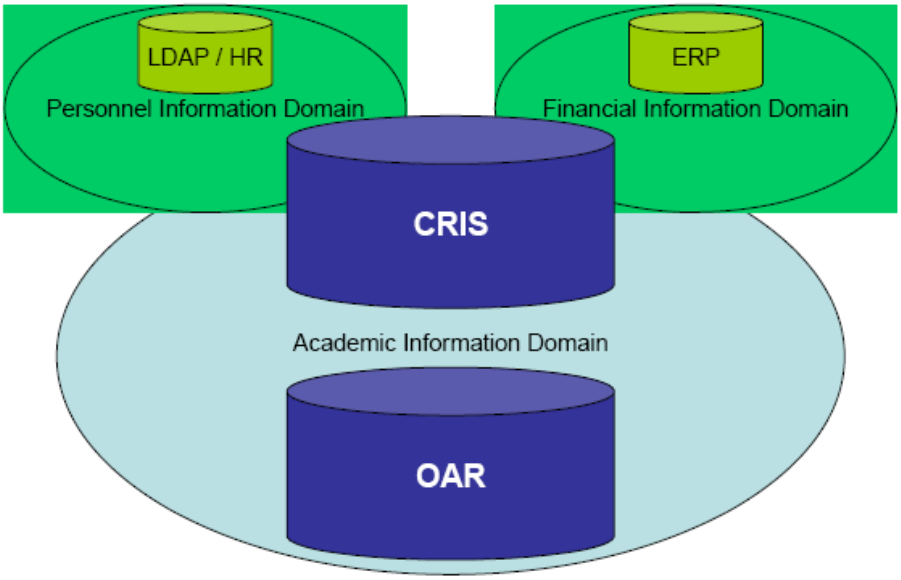


Figure 5: the Academic Information Domain

Source: Knowledge Exchange

LINKING THE DIFFERENT TYPES OF INFORMATION IN NARCIS

At the end of 2007, after most of the work of the integration of the portals NARCIS and DAREnet has been done, a new project has been started to link the information in NARCIS. This is a project carried out by the Academy, the National Library and DANS. The Academy provides information on current research. The National Library has developed a digital information archiving system, the Electronic Depot (e-depot). And DANS is responsible for preservation of research data in the humanities and social sciences. The information in NARCIS will be connected by means of unique identifiers. The connection will be realised by using the Digital Author Identifier (DAI) and the Persistent Identifier (PI).

Digital Author Identifier

In the Netherlands, but also in other countries, the CRIS and OAR of the institutions are not connected. The CRIS is maintained by the research administration, and the OAR by the library. Now one can notice a careful beginning of connecting the systems. But in NARCIS the information of the CRIS and OAR is not yet connected. To do so the first step will be to connect the two systems via the Digital Author Identifier, the DAI. Research data, publications and research descriptions may all have the name of the creator, author or project leader in common. But the names of the authors are not consistent in the different information systems. Also within one specific information system there are many author name variations.

As part of the SURFshare programme (the successor of the DARE programme)[ix] a project of importing unique names of authors/researchers in the academic CRIS (the so-called METIS) has been started last year. In the library system OCLC-PICA a thesaurus of author names (40,000) with corresponding DAI has been created within this system. All the Dutch universities, KNAW and NWO have been matching the DAI and author names in their own CRIS. This will be available at the end of 2007. The next step will be that the DAI will be implemented in the repositories of the participants. Also in the NOD, in the database of DANS (EASY) and in the E-Depot of the National Library the DAI will be imported. This is a project of the Academy, the University of Utrecht and SURFfoundation.

Because the DAI has to be harvested together with the other information the SURFshare participants had to agree on a standard in the XML. The participants have decided to use Metadata Object Description Schema (MODS).[x] This is developed by the Library of Congress' Network Development and MARC Standards Office. Unfortunately, the guidelines for the MODS are not ready yet. To take advantage of the MODS schema in an early stage, the participants will use Mini-MODS; this is a small subset of MODS.

Persistent Identifier (PI)
By using the DAI the relationship between research descriptions and research results (datasets and publications) become visible, but sustainability is not yet guaranteed. For instance: objects may shift from one repository into another or even may disappear. Or objects may be removed from a university repository, but will continue to be available in the E-depot of the national library.

To solve this problem Persistent Identifiers are necessary. Objects in repositories and datasets need PI's, so that an object is always traceable when it moves from one location to an other. It can also be used for scientific citation, since the PI can always be used to locate the object.

The PI-system that has been chosen, is the URN:NBN system of the CDNL and CENL, further developed by the German National Library. URN stands for Uniform Resource Name, and NBN for National Bibliography Number.
The National Library and DANS will be responsible for the distribution of the numbers. An example for such a number is: urn:nbn:nl-dare:a01287.

Building a DAI demonstrator
The first step is to start with the implementation of the DAI in NARCIS. It is necessary, that the harvester can harvest the metadata of the full text publication, including the DAI's. The names of the researchers in the Dutch Research Database will also be connected to a DAI. Hereafter a demonstrator will be build to show what the possibilities are after the DAI implementation in NARCIS. On the page of a researcher a link will be placed to all the publications of the researcher. It will also be possible to find all the publications of this author, if his or her name has been written in different ways (for example John Smith, J. Smith, J.S. Smith). And it will not matter if he has been working at different research institutes. This demonstrator will show the first possibilities of the DAI and will stimulate the research community to cooperate with the plans.

CONCLUSIONS AND FURTHER DEVELOPMENTS OF NARCIS

Integration of DAREnet into NARCIS
Because the Academy took over the responsibility for DAREnet it was possible to integrate DAREnet and the scientific portal NARCIS. It was advisable to integrate both portals, because like in DAREnet the repositories of the universities were harvested in NARCIS as well. The integration gave the opportunity to give an overview on research in the Netherlands. Now the portal contains information on research institutes, researchers, research activities, publications (metadata, and full text), research data (full description) and scientific news. The items (except for the news items) in the portal are searchable at the same time (one-stop-shopping), but one can also search the different information types separately.

About 30% of the full text publications in NARCIS are grey literature. Because the publications are freely accessible the development of repositories has given an enormous impulse to the retrievability of grey literature. And NARCIS makes it easy to find these publications in the Netherlands, by providing one-stop-shopping.

The Academic Information Domain
To develop the Dutch Academic Information Domain, where information from the academic current research information systems and the open archive repositories is connected, the Digital Author Identifier (DAI) and Persistent Identifiers (PI) are introduced in the NARCIS. The items in NARCIS all have the name of an author, creator or project leader in common, so the information can be interconnected using the DAI. In the Netherlands the standard of the mini-MODS will be used. To take care of preservation of the items the PI-system URN:NBN system has been chosen.

Building a DAI demonstrator
First of all the department of Research Information will build a demonstrator to show the possibilities of the DAI. It will be possible to show all the publications of an author in the context of his research activities, expertise and the research institute where he is working.

Research data
If the creators of research data will also be provided with a DAI in the e-Research repository of DANS, one can speak of a veritable Academic Information Domain, at least for the social sciences and

humanities. There are plans of the three technical universities in the Netherlands to become responsible for storing and providing permanent access to research data from the technical sciences.

Persistent Identifier
In cooperation with DANS and the National library the persistent identifiers for objects will be implemented in NARCIS to take care of the preservation of digital objects.

Usability study and user needs research
In 2006 a usability study has been carried out among Dutch information specialists. As a result some adaptations were made to the NARCIS website. Because of the integration of DAREnet and NARCIS, the portal has a new look-and-feel know. In 2008 the portal will be tested again, but know by more different users.

To find out what the needs with regard to research information of different target groups are, a user needs research will be set out next year as well. In 2002 the department has questioned four different user groups (scientists, policy makers and policy researchers, media, and industry and services (business)[xi]. In 2008 a new user needs research will be carried out.

European Dimension: DRIVER 2
Of course, these developments of interconnecting research related information take place on a European level too. The most promising project in this respect is DRIVER-II. DRIVER stands for 'Digital Repository Infrastructure Vision for European Research'[xii] The project is an initiative of 11 European countries, with the University of Athens as the co-ordinator. The project has received a grant from the European Commission (7th Framework). The KNAW and all Dutch universities are participating in DRIVER-II, united within a so-called Joint Research Unit (JRU).

Main goal of DRIVER-II is the realisation of an infrastructure that will result in a European confederation of modern repositories, in which also non-textual objects (images; datasets; presentations) can be deposited. As a spin-off, DRIVER-II aims to construct 'enhanced publications'. An enhanced publication consists of a traditional publication (for instance a report) and the objects interrelated to these publications. Here, one may think of a presentation, a dataset or a project description. The relationship with the development of the Acadamic Information Domain is clear.

References

[i] http://www.researchinformation.nl

[ii] http://www.darenet.nl

[iii] Jeffery, K.G. (2006). Open Access: An Introduction
http://www.ercim.org/publication/Ercim_News/enw64/jeffery.html

[iv] http://www.narcis.info

[v] The Role of Grey Literature in the Sciences, by Prof. Irwin Weintraub
http://library.brooklyn.cuny.edu/access/greyliter.htm

[vi] See for more information: http://www.knowledge-exchange.info/

[vii] Razum, M.; Simons, E. and Horstmann, W. (2007) Exchanging Research Information. Strand Report of the Institutional Repositories workshop, February 2007.
www.driver-repository.be/media/docs/KEIRstrandreportExchangingResearchInfoFINALFeb07.pdf

[viii] Jeffery, K.G. (2007). Technical Infrastructure and Policy Framework for Maximising the benefits from Research Output. Proceedings ELPUB2007 Conference on Electronic Publishing, Vienna, Austria, June 2007

[ix] See for more information about the SURFshare programme, coordinated by the SURFfoundation, the website: http://www.surffoundation.nl/smartsite.dws?ch=ENG&id=5463

[x] See for more information: http://www.loc.gov/standards/mods/

[xi] Koopmans, N.I., Van Meel, M. (2002) Behoefteonderzoek Onderzoekinformatie in Nederland. Nederlands Instituut voor Wetenschappelijke Informatiediensten (NIWI-KNAW), oktober 2002. 57 p. + bijl.
See for the pdf of this needs research: http://www.onderzoekinformatie.nl/nl/oi/producten/behoefteonderzoek/ (in Dutch)

[xii] See for more information http://www.driver-repository.eu/

Enhancing Visibility: Integrating Grey Literature in the SOWIPORT Information Cycle

Maximilian Stempfhuber, Philipp Schaer, and Wei Shen
GESIS-IZ Social Science Information Centre (Germany)

Abstract
Despite the fact that Grey Literature plays a key role in disseminating research results to the scientific community, it currently is not represented equally well in the established information landscape. While institutional information infrastructures supporting scholarly research (e.g. libraries or information centres) mostly focus on publications from traditional publishers or on what is published within their own institution, a parallel universe of services and servers for Grey Literature and Open Access publications has been established, which makes it hard for any user to only get an overview of what is available in his area of interest. Following a discipline-oriented approach, the newly established scientific information portal SOWIPORT not only integrates Grey Literature and Open Access publications from the social sciences with traditional ones, but also establishes support for the whole research lifecycle, where research, publication and discourse are interconnected and make use of the direct and electronic availability of these types of publications.

Introduction

With the advent of the World Wide Web and lowered technical and financial barriers that come with it, scholarly publishing in the last decade more and more shifted from print media to forms of electronic distribution. While this new technology allowed delivering a larger number of publications more timely and directly to the readers, it also put up a new challenge for those involved in collecting, documenting, managing and distributing information, namely libraries, information centres and information brokers. Not only were researchers and research institutes now able to electronically publish research outcomes or learning materials on the Web – by doing so they often circumvented the established workflow of documentation, archiving and distribution normally handled by libraries or information centres (figure 1).

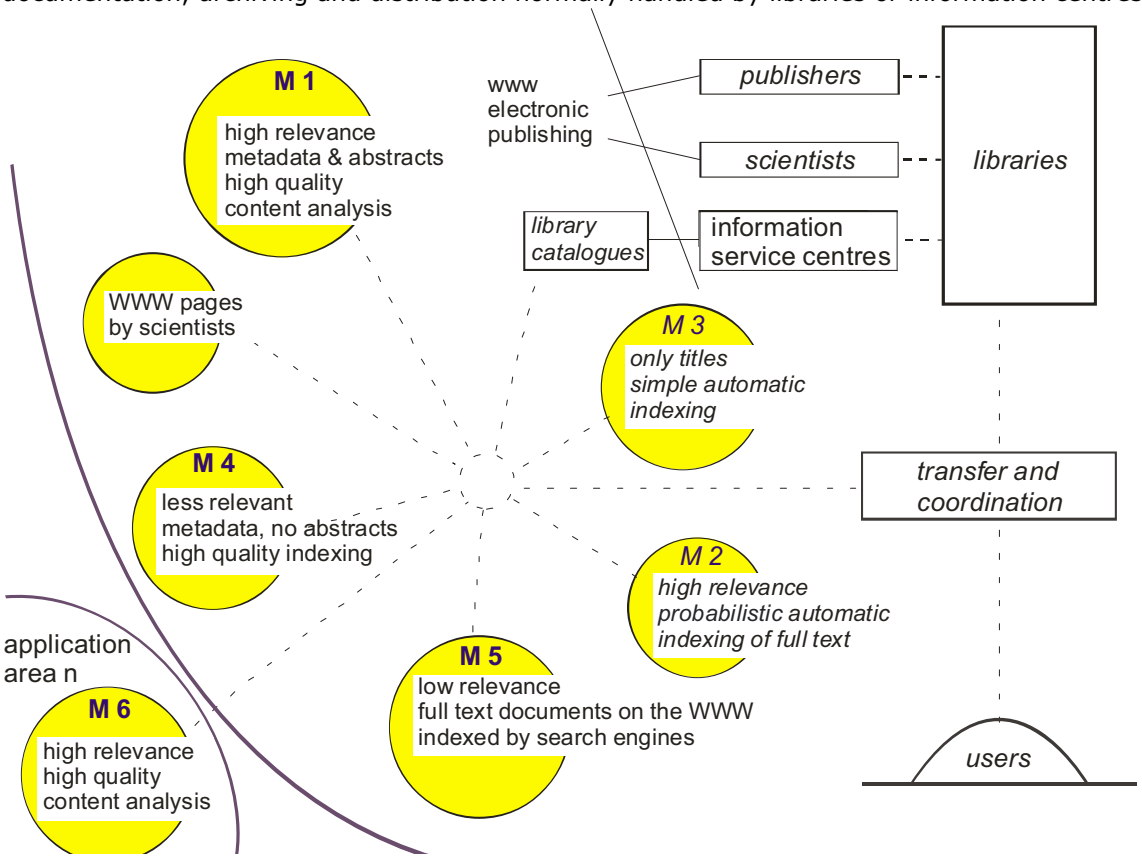


Figure 1: The changing information landscape in scholarly publishing (cf. Krause 2006, modified)

The role of libraries as agents, buying or collecting publications from commercial publishers or researchers of their own institution (e.g. technical reports) and making them available within the own institution (e.g. universities or research organizations) and to the outside world still is in effect, but there are much more sources of publications out on the Web now and many of them are not accessible through the established

publication channels. The consequences for information agents like libraries and information centres and for the user are pretty much the same:

- A constantly growing number of information sources have to be accessed to get an overview of available publications on a topic. Due to the fact that there is no central register for web sites distributing electronic publications, general web search engines like Google[i] or Microsoft Life Search[ii] are popular means of accessing Grey Literature. This in turn means that there is no kind of quality control in effect to guide users to trustworthy information. New services like Google Scholar[iii], Microsoft Live Search Academic[iv], Bielefeld Academic Search (BASE[v]) or the EU-funded DRIVER project[vi] (Digital Repository Infrastructure Vision for European Research) try to cope with this challenge in that they harvest or network digital archives with scientific content only. But still it is difficult for the user to judge the completeness of such an information service.
- The documentation (and storage) of the publications is very heterogeneous. From homepages with embedded documents without any bibliographic metadata to digital archives with standardized metadata elements (e.g. Dublin Core Metadata Element Set[vii]) freely accessible for harvesting over a technical interface (e.g. OAI-PMH[viii], the Open Archives Initiative's Protocol for Metadata Harvesting), every mix of bibliographic information, content indexing, classification and abstracts can be found. But not only is the structure of the metadata very heterogeneous, also the semantics do vary because of the use of different controlled (classifications and thesauri) and uncontrolled (author keywords) vocabularies for classification and content indexing. Together with the different natural languages used in the metadata, this leads to a very complex situation for indexing and retrieval and puts up challenges for the user when trying to search several sites for relevant publications and also for information providers when trying to harvest metadata and integrating them into a new service.
- The scope of traditional publications, journals and series is most of the time focussed on a specific domain or discipline, making it relatively easy for a user interested in a dedicated topic to identify relevant sources of information. For Grey Literature and Open Access publications, a large number of institutional repositories have been established, many of them focussing on the publication output of a whole organization rather on the discipline the individual publication belongs to. This, in combination with the heterogeneity of metadata across repositories, makes it sometimes very difficult to select publications by discipline even from a single repository. When integrating content from multiple repositories, the effort for filtering information by discipline is even higher.
- Grey Literature and Open Access publications are still seen as second-class publications in regard to scientific career and promotion across all disciplines, despite the fact that many of these publications went through the same type of peer review as traditional publications (cf. JISC/OSI 2004, Swan&Brown 2004, Rowlands et al. 2004, Rowlands&Nicholas 2005). One of the reasons for this might be the fact that many of these publications are not documented in the established and well known reference databases and library catalogues, which limits their visibility and credibility.

The challenges raised by the Web's infrastructure for electronic publishing must also be seen in the light of users' expectations towards (scientific) information services. Several surveys on user satisfaction and demands towards library information services (cf. RSLG 2002, Poll 2004) indicate, that especially the discipline-oriented view on information and the integration of all available types of information is what attracts them:

- Domain-specific organization: Users demand for domain-specific organization of information, so they have a single point of access to all potentially relevant materials from their discipline. There should be deep content indexing, allowing for precise search and sophisticated search strategies.
- Interdisciplinary connections: At the same time, users see the need for crossing the borders between domains and disciplines. Provisions should be made for letting users expand their queries beyond the scope of the information service they are currently accessing.
- Broad scope of information: To satisfy most information needs, all relevant types of information should be made accessible at one place: Primary data (e.g. surveys and other raw data), secondary data (e.g. aggregate and time series data), and references to literature, research projects, internet resources, experts, networks, software etc.
- Quality of service: To help users satisfy their information needs, results should be limited to a manageable amount of relevant information (no information overflow), noise and low quality content should be avoided and all materials should be instantly accessible right from the desktop ("now-or-never" paradigm) or they might not be used at all.
- Informal communication: Besides the need for accessing information for carrying out research there is also a strong need for communication, like discussing results with colleagues in discussion boards or exchanging information and papers via e-mail. This informal communication (in contrast to the formal communication by means of traditional publications) is of more dynamic nature, subject to small and even closed groups of researches (invisible colleges), and currently often separated from the general information services offered to a discipline. The growing importance of this type of interaction for the generation of new ideas and even of new

types of research and publications (Harnad, 1991; Nentwich, 2003) makes it inevitable to integrate it more seamlessly with traditional offerings.

The SOWIPORT Information Cycle

With the new social science information portal SOWIPORT[ix], GESIS[x] (German Social Science Infrastructure Services) and its partner organizations try to overcome the problems outlined above and to satisfy the needs and demands of scientific users. The main goals of SOWIPORT, which is based on the results of several research projects funded by Deutsche Forschungsgemeinschaft (DFG) and German Federal Ministry for Education and Research (BMBF), are to provide central access to social science information from distributed sources, to tightly integrate this heterogeneous information and to enable new ways of scholarly publishing and communication.

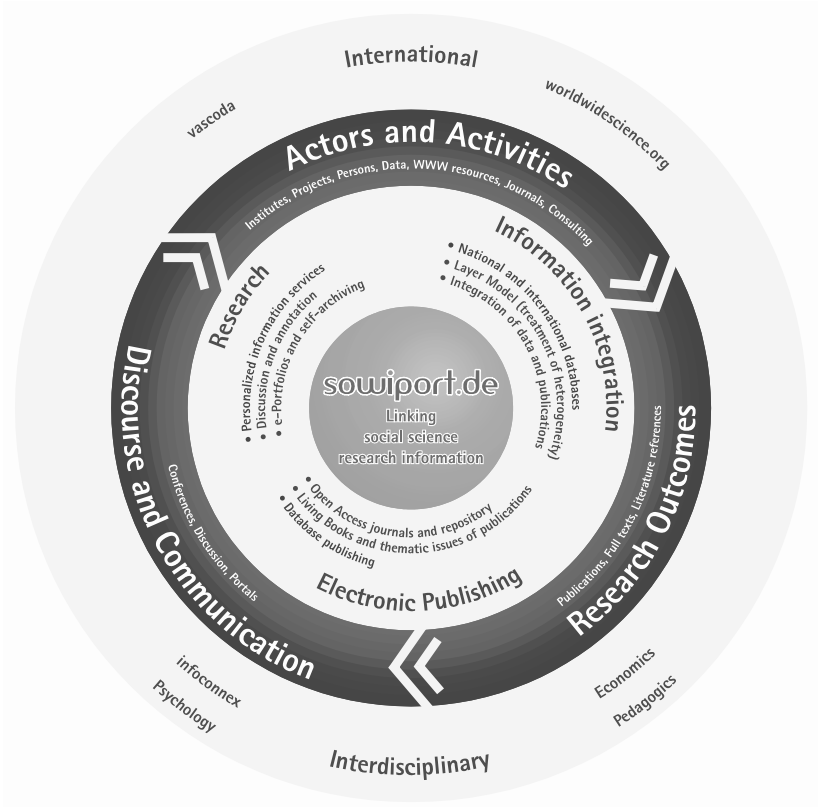


Figure 2: The SOWIPORT information cycle

The concept of SOWIPORT is based on an information cycle (figure 2) which – in contrast to other models in the context of libraries or research information systems – does not focus on different categories of publications or data (e.g. primary, secondary and tertiary sources), or on the demand for or creation of specific types of data or publications at different stages during research activities. Rather it divides scientific information and services for researchers into three groups, each representing a different activity associated with being part of a scientific community:

- Actors and activities (making up the community): This group contains all structural information about a domain or discipline, like the persons and institutes doing research, the projects they are carrying out, the services available to researchers in this discipline (e.g. consulting for questionnaire design, archiving of data, computing resources), the most important journals or publishers, available datasets etc. The focus here is on information which gives an overview of the domain and helps researchers in networking and organizing their research.
- Research outcomes (publishing results): This group of information currently includes publications as the most important type of research outcomes. An important aspect here is that published research outcomes are approved by the scientific community, normally through some kind of reviewing process. Since more and funding organizations are discussing whether to require researchers to also publish their primary data along with their publications, the diversity of research outcomes might increase in the future.
- Discourse and communication (interacting with colleagues): This group consists of information about conferences, discussion boards and mailing lists, and of thematically focussed information services like current awareness services or special interest portals which might include editorials and selected information from "actors and activities" and "research outcomes".

Embedding Grey Literature and Open Access publications in SOWIPORT

These three groups of information are tightly integrated by several services provided to users in the context of SOWIPORT. From the perspective of integrated information access, the treatment of the heterogeneous metadata for entities like publications, research projects, data and institution profiles is a necessary precondition to overcome the current situation, where users have to access heterogeneous data sources either sequentially (requiring them to re-formulate their query several times) or the data is simply pooled and it is left to the user to come up with all potentially relevant keywords for his information need. In SOWIPORT, cross-concordances between the multi-lingual thesauri of currently around 15 databases are used to automatically translate a user's query terms to the controlled vocabularies used to index these databases. Cross-concordances are bi-lateral mappings between pairs of thesauri and consist of relations like equivalence, broader terms or narrower terms between the descriptors of the thesauri. This semantic technology also allows us to translate concepts from the social sciences to the terminology used in other disciplines and therefore to expand a user's search across different disciplines, e.g. psychology and pedagogics. The portal infoconnex[xi] already contains this type of interdisciplinary search which also will be integrated in a future version of SOWIPORT. The technology will also be used for searching in harvested metadata from institutional repositories (Grey Literature and Open Access publications), in which often keywords instead of controlled vocabularies are used for content indexing (often done by the authors themselves instead of librarians). By expanding a user's original query using cross-concordances and using the expanded query for searching in harvested metadata, the recall can be improved and users will get results for queries which otherwise might have yielded no results in these collections because of the broad variety of keywords used there in comparison to the standardized vocabularies of library catalogues and reference databases.

The concept of electronic publishing connects research outcomes with scholarly discourse and communication in SOWIPORT. One central component here is the Social Science Open Access Repository (SSOAR[xii]), which will be integrated in SOWIPORT in early 2008. SSOAR is a project funded by Deutsche Forschungsgemeinschaft and carried out by GESIS-IZ and Freie Universität Berlin with the goal not only to set up a repository for freely available publications, but also to integrate the repository with the established information infrastructures like reference databases and library catalogues. SSOAR seeks to establish a workflow where materials uploaded by authors are first checked for domain-specificity and technical quality by peers from the community and then – after passing additional formal criteria – are documented in reference databases at the same level traditional publications from journals are documented (full set of bibliographic information, content indexing with thesauri, creating abstracts, linking to full text etc.). For the duration of the project, Grey Literature and Open Access publications with at least one German author or editor are processed this way to gain some data about the amount of resources needed for documenting these types of literature. For publications beyond this well-defined scope, the help of the scientific community will be required and there are already signs that some libraries are interested in supporting the project. Hope is, that not only visibility and accessibility of Grey and Open Access Literature will increase by improving the level of documentation and including it in reference databases, but also that authors and institutes without own repositories will be motivated to upload their publications to SSOAR. For long-term preservation, negotiations with the German National Library are on their way.

To foster discourse and communication between the actors of a discipline, SOWIPORT tries to include several ways to let researchers take part in the creation of content. One example are personal homepages (currently under development) where researchers can publish their CV, scientific profile, publications and research projects they are working on. Most of the information is already contained in SOWIPORT (with a focus on German-speaking countries), so that researchers do not have to enter all the information from scratch but can link from their homepages to the content of SOWIPORT's databases. They will for example be able to search in all of the SOWIPORT databases for their publications and just select them for inclusion in their personal homepage. At the same time they will be able to upload the full text of their publications into the SSOAR repository and add all publications currently not covered by any of the databases. The general hypothesis is, that by overcoming the currently predominant separation of personal homepages, institutional repositories and established databases, and integrating them within a single portal, the systems researchers normally use for acquiring information would merge with the systems they use for distributing their research outcomes or information about themselves. Self archiving of Grey and Open Access literature by authors would automatically improve the timeliness of databases and library catalogues, the inclusion of these materials in those systems would at the same time improve their visibility.

Using the potential of digital publications

Grey Literature and Open Access publications have due to their machine-readable format and electronic availability the potential to become the basis of new types of scientific publications in which scientific discourse and collaboration play a central role (cf. Nentwich, 2003). Within SOWIPORT, prototypes of such new forms of publications are under development, with in-context discussion for journal articles being the first.

The basic idea for in-context discussion started with the observation that in many disciplines, online discussion about scientific publications is by far less frequent compared to online discussion in news portals, where in some cases several hundreds of comments can be read just within hours after a news item is published. Beside differences between disciplines in the culture of public scientific discourse some interesting formal differences between scientific articles and news can be seen, with the big difference in text length being the most obvious one. From a human factors point of view, the separation of the context of a discussion (the article or news item displayed in the user's browser) and the actual discussion (which takes place in some other browser window or web page) puts a huge cognitive load on the user which will increase with the length of the article, the complexity of its content and the distance (measured in pages, windows or mouse clicks) between the article and the online discussion. The longer the article, the more difficult it gets to refer from one point in the discussion to the precise location in the article which is currently being talked about, and the more mouse clicks necessary to switch back and forth between article and discussion the bigger the load on a user's short term memory to either remember what was written in the article (when reviewing or writing comments) or what was said in the discussion (when comparing statements made by other readers with the content of the actual article). The design goal for online discussion about electronic publications in SOWIPORT therefore was to bridge the gap between the two media – the article and the discussion – and to combine them into a new, single media. At the same time, the formal communication represented by the published article should be merged with the non-formal communication as used in online discussion.

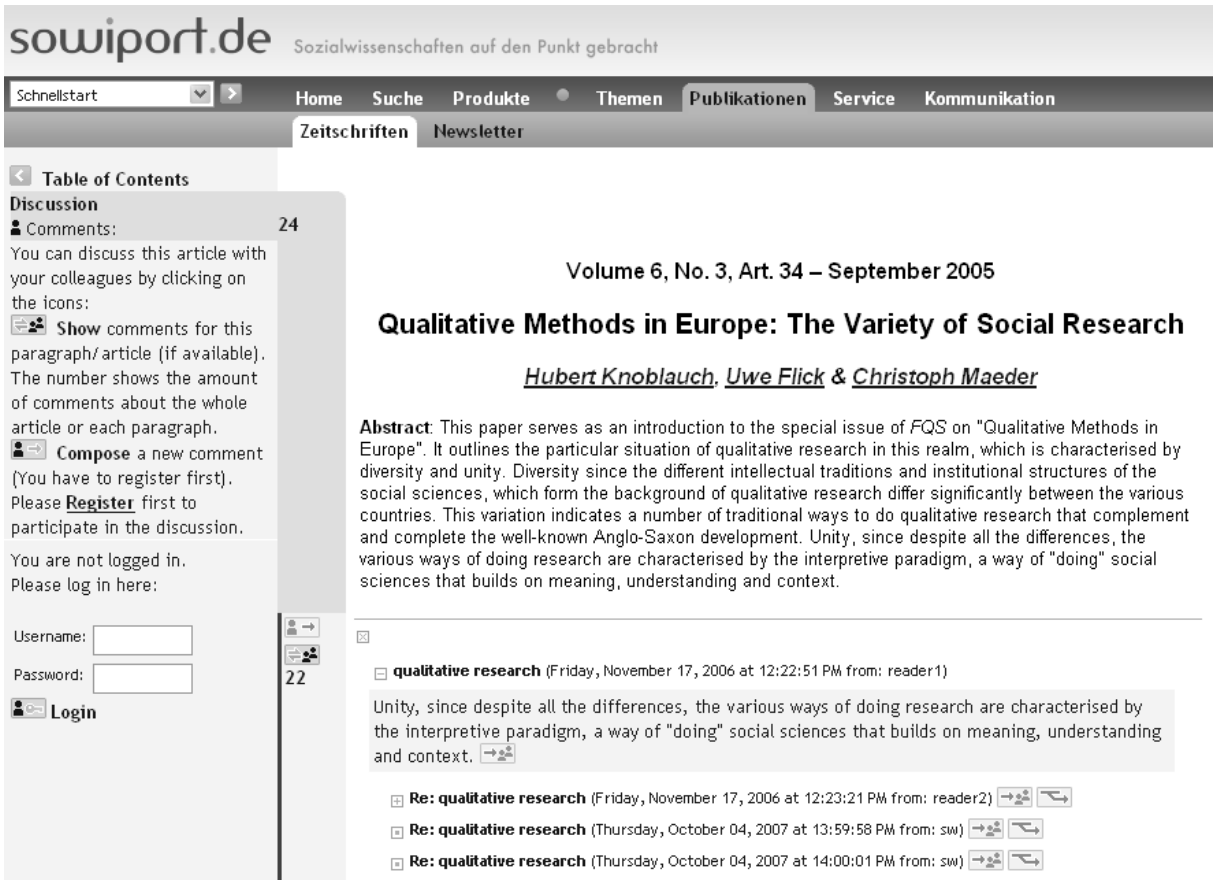


Figure 3: Integration of online discussion into a published article

Figure 3 shows how the integration has been realized in SOWIPORT so far. The left margin of the article display indicates with an icon at each paragraph that a reader may discuss this section of the article after registering in the portal or that there already exist comments for the specific paragraph of the article (indicated by an icon and the number of existing comments). The paragraph as granularity for discussion has been chosen because of the fact that publications may be multi-language and translation will often modify sentence structure to allow for better style of the translation. Therefore it can not generally be assumed that e.g. sentence number 14 in the English version will be the literal translation of sentence number 14 of the Spanish original. The comments are dynamically displayed just between the two adjacent paragraphs without any reload of the page and can be hidden again in much the same way, which gives users the impression of a static display of the article which can be enriched with information about the ongoing discussion at any time, thus putting the user into full control of the time and amount of the information displayed.

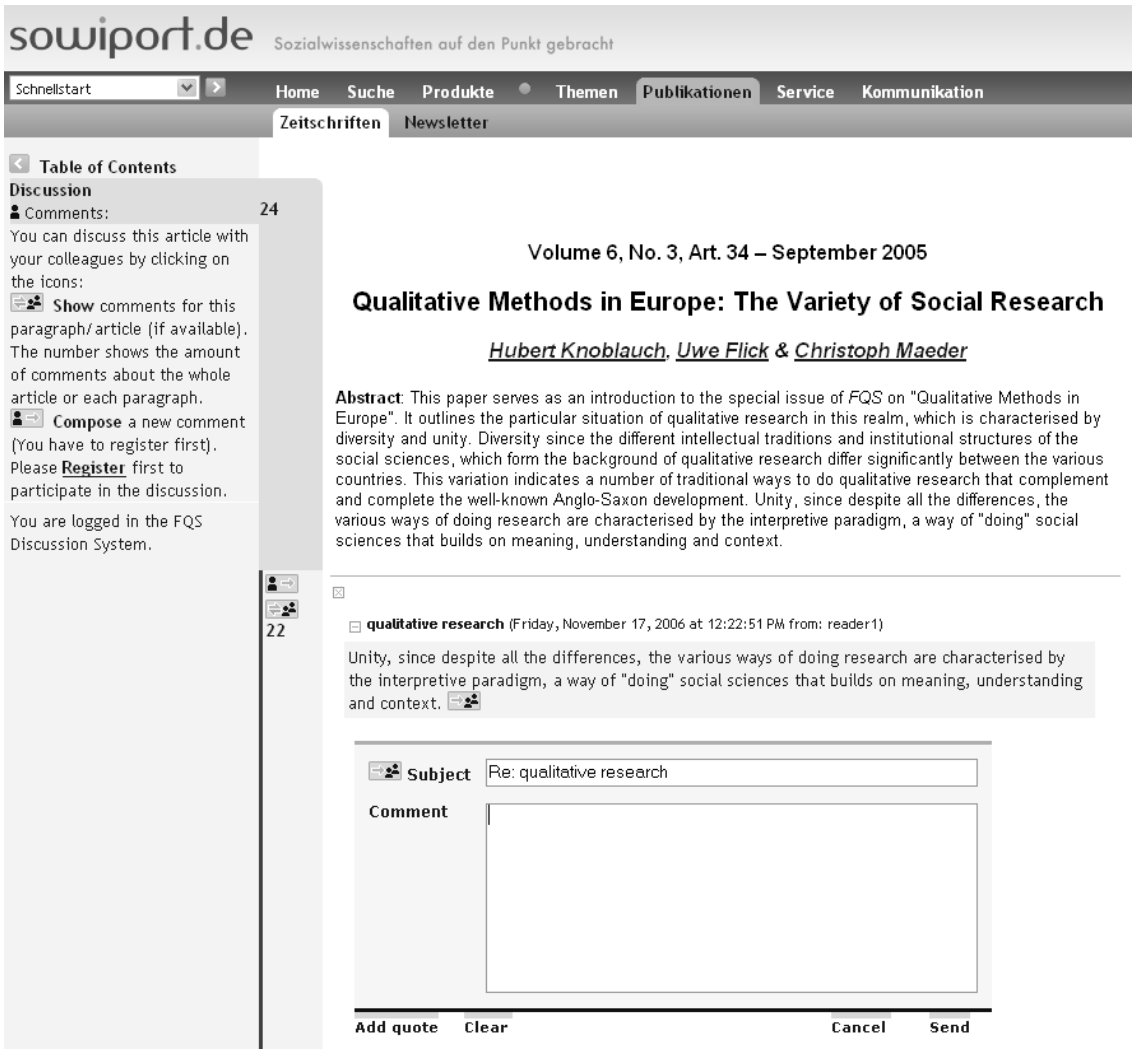


Figure 4: Following up with new comments

In the same way as comments are displayed between paragraphs, an entry form is dynamically inserted into the article, letting users create new comments while still being able to read the comments they are answering to or the paragraph they want to discuss about (figure 4). All the non-formal communication (the discussion) takes place within the frame of the formal communication (the published article) thus making the discussion an integral part of the article (and therefore referable) without destroying the integrity of the surrounding article.

Conclusion

The social science information portal SOWIPORT seeks to integrate Grey Literature and Open Access publications much more with the established information products like reference databases or library catalogues than it is currently the case in many disciplines. The approaches used to realize this are motivated by a more functional or action-oriented view to the information cycle, bridging different groups of information by functional features like integrated search, electronic publishing and scientific discourse. By integrating self-archiving of publications into SOWIPORT and connecting these often non-traditional and hard to find types of publications with the established reference systems for the discipline, we hope to raise their visibility to the community and to create an incentive for authors to make them available. The availability of many different types of information on one platform holds the potential to create new information products or new types of publications by the users themselves. The in-context discussion in electronic publications is a first step which will be evaluated concerning usability and acceptance by the users. As an enhancement to this feature we are planning to integrate references into the users' comments, which will allow to directly point to other publications, data sets or entities contained in the SOWIPORT databases.

References

Harnad, S. (1991): 'Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge', The Public Access Computer Systems Review, vol. 1, no. 2, pp. 39-53.

JISC/OSI (2004): JISC/OSI Journal Authors Survey, February 2004. (http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf)

Krause, J. (2006): 'Shell Model, Semantic Web and Web Information Retrieval', in I. Harms, H.-D. Luckhardt & H. W. Giessen (eds.): Information und Sprache. Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern. Festschrift für Harald H. Zimmermann. K. G. Saur, München.

Nentwich, M. (2003): Cyberscience. Research in the Age of the Internet. Austrian Academy of Sciences Press, Vienna.

Poll, R. (2004): 'Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung, Teil 1: Informationsverhalten und Informationsbedarf der Wissenschaft', Zeitschrift für Bibliothekswesen und Bibliographie, no. 51 (2004), pp. 59 - 75.
Rowlands, Ian; Nicholas, Dave & Huntingdon, Paul (2004): Scholarly Communication in the digital environment: What do authors want? - Findings of an international survey of author opinion: project report. (http://www.ucl.ac.uk/ciber/ciber-pa-report.pdf)

Rowlands, Ian & Nicholas, Dave (2005): New journal publishing models: an international survey of senior researchers. (http://www.slais.ucl.ac.uk/papers/dni-20050925.pdf)

RSLG (2002): Researchers' Use of Libraries and other Information Sources: Current Patterns and Future Trends. Final Report. Education for Change Ltd., SIRU, University of Brighton & The Research Partnership (http://www.rslg.ac.uk/research/libuse/).

Swan, Alma and Brown, Sheridan (2004): Authors and open access publishing. In: Learned Publishing (2004) 17, 219–224. (http://cogprints.org/4123/1/Authors_and_open_access_publishing.pdf)

---

[i] http://www.google.com
[ii] http://www.live.com
[iii] http://scholar.google.com
[iv] http://academic.live.com
[v] http://www.base-search.net
[vi] http://www.driver-repository.eu
[vii] http://dublincore.org/
[viii] http://www.openarchives.org
[ix] http://www.sowiport.de
[x] http://www.gesis.org
[xi] http://www.sowiport.de
[xii] http://www.ssoar.info

Nancy STYLE

a tool to improve the production of Grey Literature

What is Nancy Style?

It is the informal name given to the

Guidelines for the production of scientific and technical reports: How to write and distribute Grey Literature

formally presented by the Istituto Superiore di Sanità (Rome, Italy) during the 7th International Conference on Grey Literature held in Nancy (France) in December 2005.

Who can use this tool?

Authors and GL producers in their mutual task of creating and distributing accurate, clear, easily accessible reports in different fields.

Which goal?

Permit an independent and correct production of institutional reports in the respect of the basic editorial principles.

Which language?

The original version is in English. Its translations are in:

- Italian (by **Istituto Superiore di Sanità**)
- French (by **Institut de l'Information Scientifique et Technique**)
- German (by Technische InformationsBibliothek/UniversitätsBibliothek)
- Spanish (by Universidad de Salamanca, *in preparation*)

Where can you get it?

All the versions are available from the official site of the GLISC: www.glisc.info.

What is it about?

- Ethical principles related to the process of evaluating, improving, and making available reports, and the relationships between GL producers and authors.
- Technical aspects of preparing and submitting reports.

What is GLISC?

The group approving these Guidelines is formally defined as Grey Literature International Steering Committee, composed of:

Istituto Superiore di Sanità (ISS), Italy

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

Grey Literature Network Service (GreyNet), The Netherlands

For information:
www.glisc.info
secretariat@glisc.info

Paola De Castro and Sandra Salinetti
(Istituto Superiore di Sanità, Rome – Italy) 2007

Assessment and improvement of a corporate research information system

Maria Castriotta
ISPESL – Departiment of Documentation, Information and Training, (Italy)

Daniela Luzi
CNR – National Institute of Population Research and Social Policies (Italy)

Mariarosaria Manco
LINK s.r.l. (Italy)

Introduction

In 2003 the Italian National Institute for Occupational Prevention and Safety (ISPESL) developed a Research Information System in the field of Occupational Safety and Health (RisOSH) through a project in collaboration with the Italian National Research Council (CNR). RisOSH became operational in 2005 and it is now embedded in the Institute webpage dedicated to Research Activities (www.ispesl.it/ricercheOSH/ext/). Its double-fold interface supplies information on conditions and modalities to participate in project calls, permits the electronic input of research results, and allows users to search for ongoing and finalized projects. It can also constitute an access point for internal users, through a reserved area, to directly contribute to the collection and validation of data within the project lifecycle.

The webpage has shown a very positive trend in terms of access, counting a total of more than 15000 connections per year, with high numbers of users performing searches in the database with the use of descriptors (> 4000/year). This trend is very encouraging because it shows the interest of users in looking for qualitative information; further access data regarding the quantity of pages exploited *browsing* through the results (>22000/year) and for printing the full record (>1200/year), confirm the users' need of an exhaustive information on research projects.

The aim of the present study is the evaluation of the RisOSH system after more than two years of maintenance, in particular regarding the number of projects concluded and the quality of the results obtained. The system was designed and developed describing in detail the project lifecycle based on ISPESL organisation and giving a particular emphasis on the phase of collection of projects results. This is in our opinion the principal intent of a Current Research Information System (CRIS): providing the users not only with information on project aims, rationales and participating organisations, but giving them a comprehensive picture of the achievements of the projects, in terms of project result descriptions, availability of the final project deliverable, as well as product descriptions and publications. In this way the multiple purposes of a research information system can be obtained: diffusion of research results to the scientific community; providing the top-management and/or policy makers with information on scientific field to be granted; building an instrument that is able to improve technological transfer.

Following this line, an assessment was carried out after the development and publication of a new template for the web collection of project results, containing new editing areas that describe, more exhaustively than the previous form, the outcomes of research projects. The paper describes the results of the evaluation analysing on one hand the effectiveness of the new electronic template, and on the other, the quality and quantity of the data gathered through this tool. A further step was put into action, consisting in a plan of contacts within the Institute scientific community, aimed at making researchers more aware of the system use and involving them in providing missing information or in integrating or amending the available data. The assessment resulted in the identification of the system technological and organisational strengths and weaknesses, confirming, even within the limits of a scarce commitment of the Institute as a whole, the important role played by RisOSH in the construction of an ISPESL institutional repository, with its features of transferability and interoperability with other intramural and extramural information systems.

1. Objectives and methodology

The collection of project results has been considered by many authors a real challenge [CRIS2006, EUR2004, JEF2002] for CRISs, as in several cases an initial detailed description of project objectives, work packages and information on the institutions involved does not often match with an adequate reporting of projects results and products. Moreover, deliverables and/or publications produced within the project are rarely available. This causes a loss of information for both researchers and research policy makers, who could take advantage of these results to develop new projects and foster activities in specific promising

research fields. This information loss also makes the technological transfer to potential users far more complicated. For this reason, we consider the collection of projects results as an outstanding feature of CRISs quality.

Therefore, our evaluation activity was based upon the project result descriptions and had the aim of:

- analysing RisOSH information content both from a quantitative and qualitative point of view;
- verifying the efficacy of the information system, considering in particular the new electronic template developed to collect the project results.

Our general intent was to improve the communication flow of the information producers in order to increase the number and enhance the quality of the collection of project results, identifying the best practice for a direct involvement of the ISPESL's scientific referees of projects as well as a greater commitment of external project assignees.

To reach this goal we first analysed the data contained in the system before we started the evaluation activity, i.e. before February 2007 and identified the projects with missing result descriptions. Then we contacted ISPESL's scientific referees via e-mail and asked them to update information. Afterwards, we analysed the answers received, focusing on the quality of project results, and in particular:

- Consistency of result descriptions (i.e. Grey information)
- Availability of deliverables (i.e. Grey Literature)
- Availability and description of products (i.e. Grey information)
- Publications produced during the project.

2. Main features of the RisOSH system

Before dealing with the evaluation process carried out in the present study, an overview of the main issues considered in the design of RisOSH is presented, showing the solutions adopted in its first planning and development. A preliminary analysis of the information systems for the collection of research projects identified the following main features: the existing standards; the classification systems available in the specific sector; the information producers, tightly connected with the organizational structure; the modalities of diffusion to be activated.

The analysis of the existing standards was carried out with the aim of identifying a standard model allowing the development of open systems for the integration with different databases; such analysis brought to the identification of the database schema: the CERIF2000 standard (Common European Research Information Format) [CER2000] was adopted, which gives a detailed description of R&D projects and is used in CRISs. This makes the system interoperable with other CRISs.

As to the choice of classification systems for the indexing of the projects, the Classification of Economic Activities [ATECO2002] was identified to facilitate the technological transfer and the *Occupational Safety and Health Agency OSHA Thesaurus* was chosen to describe the aspects related to safety and health at work. Both classifications have multilingual versions and this facilitates the exchange of information especially at an international level.

A thorough analysis was also carried out on the information sources produced in the course of the realization of research projects, in order to facilitate the system feeding and updating as soon as the information is produced; this brought to the identification of the information producers, who are involved in the project lifecycle, from the phase of project proposal to the collection and evaluation of project results. Through the description of the project lifecycle and of the related information flow, it was possible to identify the actors, task performed, time scheduling and exchange of information and documents of the entire process.

Tightly connected with the information sources is the organizational structure, analysed in order to identify possible bottlenecks as well as parts of the process which need the development of specific system functionalities that will help the organization model becoming more effective. For this reason, RisOSH reconstructs a workspace, both in the functionality and in the interface design, which is very similar to that used before its introduction. The similarity with the daily workflow has the advantage of facilitating user interaction with the new information tool, and at the same time, of increasing motivation to use it. The backoffice structure of RisOSH is based on the workflow methodology.

Figure 1 represents - in the sequence of the menus as well as graphically - the different phases of the process: the elaboration of the proposal ("Proposta"), the editing and publication of the activity plan ("Piano") and of the call for proposals ("Bando"), the project assignment ("Assegnazione"), and the collection of the research results ("Risultati"). The only menu, which can be accessed at any phase of the process it that related to the assignment of the descriptors which classify the project ("Indici"): indexing can either be performed in the phase of the proposal elaboration, and/or updated by the external research institution once the project is concluded.

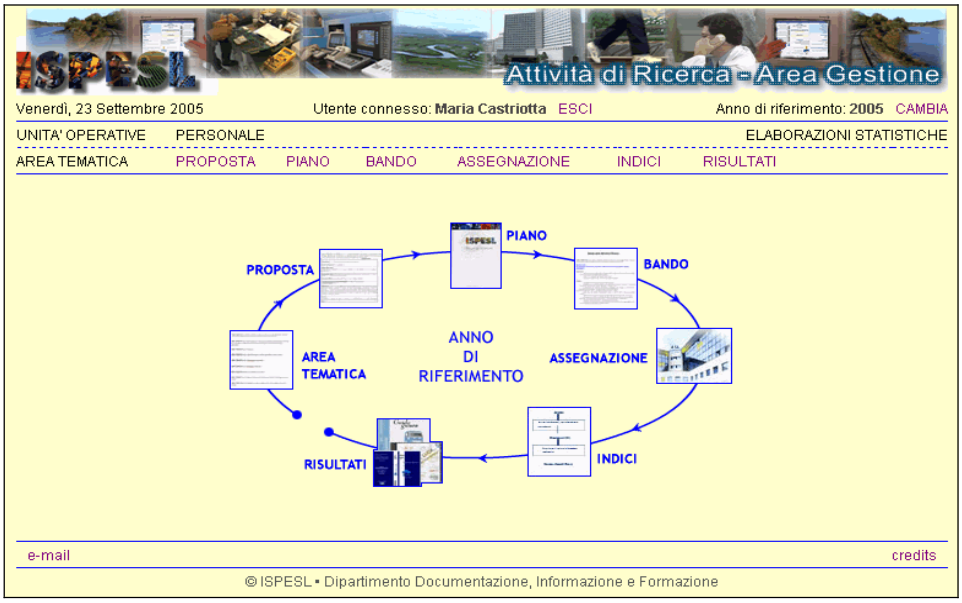


Figure 1 – Main menus of the RisOSH system

Concerning the modalities of diffusion of research results, a friendly web interface was created with the incorporation of RisOSH. The system can therefore be accessed from the ISPESL's homepage "Research Activities", which allows retrieval of projects using the following options: the browsing of NACE/ATECO codes; the browsing of OSHA descriptors; a simple textual search, a guided search (Fig. 2).

Figure 2 – Webpage "Research Activities" within ISPESL website

The system also provides the users with further information on the modalities of participation to projects calls, on the list of projects assignees that will have to sign a contract with ISPESL, as well as with a predefined template that gathers the project results ("Scheda di sintesi"), to be filled once a project is finalized.

3. Improvement and assessment of the system

The number of projects funded by ISPESL to external institutions, that is the object of the RisOSH system up to now, has steadily decreased in recent years because of progressive reduction of general research funding. Moreover, a trend is observed toward the concentration of money on a reduced number of bigger projects, lasting for many years.

RisOSH currently contains more than 600 descriptions of research projects funded by ISPESL in the period 2000 to 2004. The descriptions of projects, most of which started respectively in years 2002, 2003, 2004 and 2006, were inserted during the system development taking the information from official documents, such as the ISPESL's Activity Plan and the publication of the list of assigned projects to external organisations that carry out the research in collaboration with ISPESL. The input of data on project objectives, rationale, project staff and organisation was done centrally in order to test the system functionalities as well as to provide users with core information that could facilitate its future updating.

However, coherently with the technological approach adopted, that privileged a distributed input of data by the information producers, two steps were planned in the finalization of the project: the collection of results through an online template, to be filled in by the project assignee; the assessment of the results by the ISPESL scientific referee, who, once validated the data, could be able to directly publish the project results on the internet, deciding also whether publishing the full-text of the final deliverable.

Moreover, anticipating the development of the RisOSH system, ISPESL administrative offices accepted to include among the research contract clauses, the request of compilation, by the project assignee, of an online template containing a synthesis of research results, together with the hard copy of the final full deliverable. Therefore, a first template was created for the acquisition of projects results and made available on the Institute's website since 2002.

With the aim of improving the number and quality of project result collection, in 2005 a new more complete version of the template was developed and published on the internet in substitution of the former version. Such template is a subset of the database schema and the data collected are at the moment saved in a subfile for their processing at a centralized level; the system is however ready for a shared input.

In the development of the new version of the template, particular emphasis was given to the collection of data regarding products and publications produced in carrying out the project. The request of downloading the full-text of the project deliverable was also a new important feature of the template.

Figure 3 shows the new template for the acquisition of research results, containing six editing areas: general data ("Dati Generali"); final deliverable ("Elaborato finale"); classification ("Classificazione"); products ("Prodotti"); publications ("Pubblicazioni"); technological transfer ("Trasferimento tecnologico").

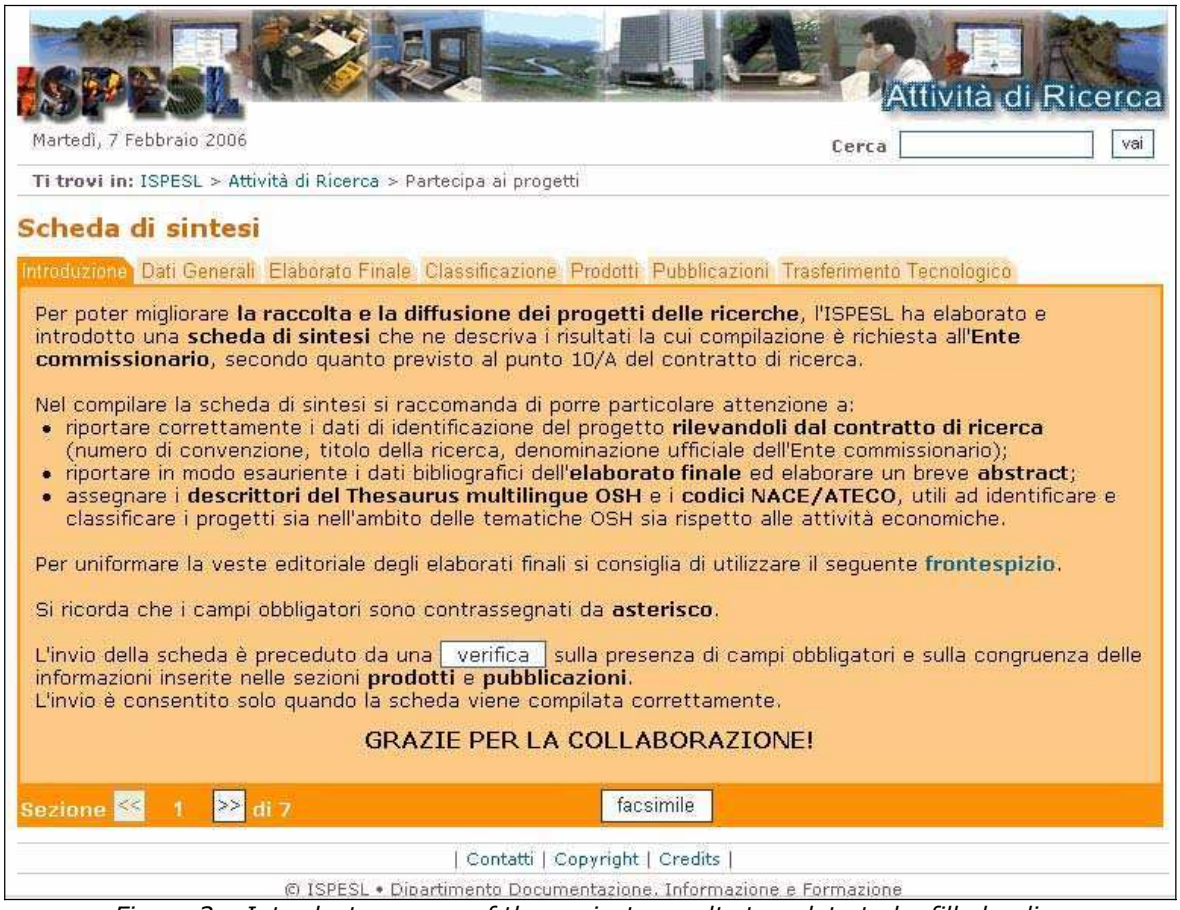


Figure 3 – Introductory page of the projects results template to be filled online

The final deliverable area collects the bibliographic description of the final report, as well as an abstract of the project results. These data are in English and in Italian; actually, the database schema supports multilingual versions and this allows the implementation of an English version of RisOSH. For the

completion of the information chain represented by the project lifecycle, the integration with the ISPESL Grey Literature Database [LUZ2005] was also realized, with the aim to use in an efficient way the data coming from different channels. This integration is based on a software that transfers the final report bibliographic data into the Grey Literature Database after checking and matching it.

The classification area collects the descriptors that the assignee has to identify in the OSHA Thesaurus to classify his project. The system helps the user in choosing descriptors: selecting the thesaurus button, the main categories are shown and searchable. If a descriptor is selected, it is copied in the project template. A similar procedure is present in the technological transfer area, where the project is classified through NACE codes [ATECO2002]. This classification is of outstanding importance for the types of projects carried out in the OSH sector, that mainly cover research applied to productive activities.

Finally, two sections are dedicated to a kind of data that we consider as fundamental for a qualitative description of projects results: the sections collect descriptions of guidelines, databases, websites, training modules, etc. produced as realizations of the project, as well as articles and papers published in reference to the project results. A lot of aids have been created for the right compilation of the template, together with checks and validation of data before the online submission.

4. Results of the assessment

In the first part of our assessment activity, we analysed the RisOSH information content. The data contained in RisOSH until February 2007 are shown in figure 4.

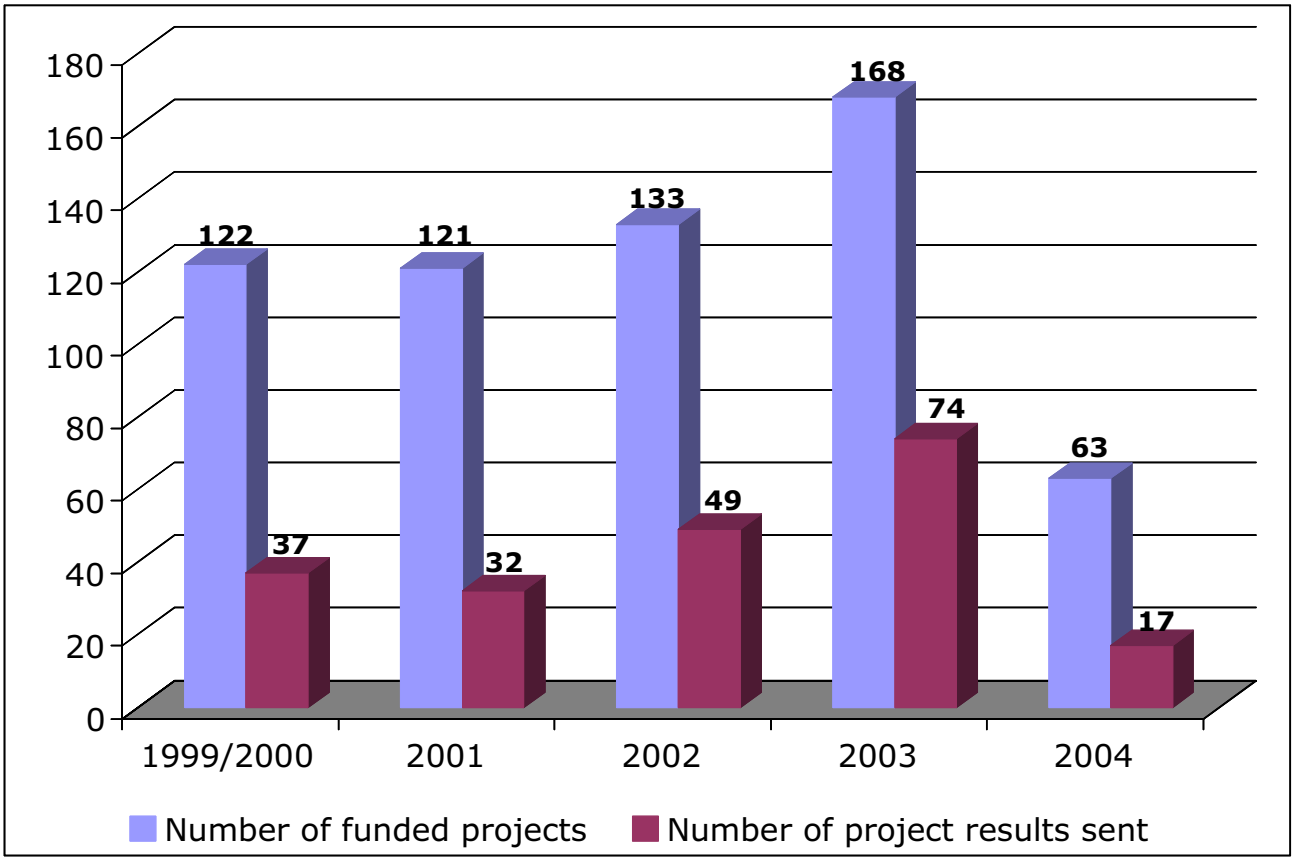


Fig. 4 - RisOSH data until Feb. 2007 distributed by year of project funded & description of project results

Since 1999 ISPESL funded 607 projects; 34.4% of them (209) contains information about project results. It is not a high percentage, but we can observe a slight increase which is especially remarkable for the projects funded in 2003, whose result descriptions reach 44% (74 out of 168); while the majority of projects funded in 2004 are not concluded yet. The percentage is also not quite high because the clause included in the research contract (online transmission of project results) is not really checked in the phase of assessment of the project results, nor is the necessary condition to get the final project funding.

A more promising picture of the result descriptions sent by the assignees is given by the quality of the information reported in the electronic format. Only a minority of these formats could not be inserted because of lack and/or inconsistency of the data reported. Moreover, the result descriptions collected with the revised formats allows the assignees to report bibliographic data of publications produced during the projects (30% of the project funded in 2003 and 47% of the projects funded in 2004) and describe the

products development, such as software, databases, guidelines, etc (44% of the projects funded in 2003 and 23% of the projects funded in 2004).

It is also interesting to note that a very high percentage (more than 85%) of the assignees, download the full text of the final deliverable using the facility provided by the new electronic format. In this way the repository of ISPESL's GL database is going to increase.

In the second phase of our evaluation activity we prepared the e-mails to be sent to ISPESL's referees soliciting them to review and update information on project results using the available electronic format. We received about 20% of answers, which allowed us to update RisOSH as well as to receive important information about changes in the organisation of the project and/or about projects that has been prolonged. The number of answers received was lower than our expectations, but the process resulted very positive in terms of new contacts, suggestions and appreciation.

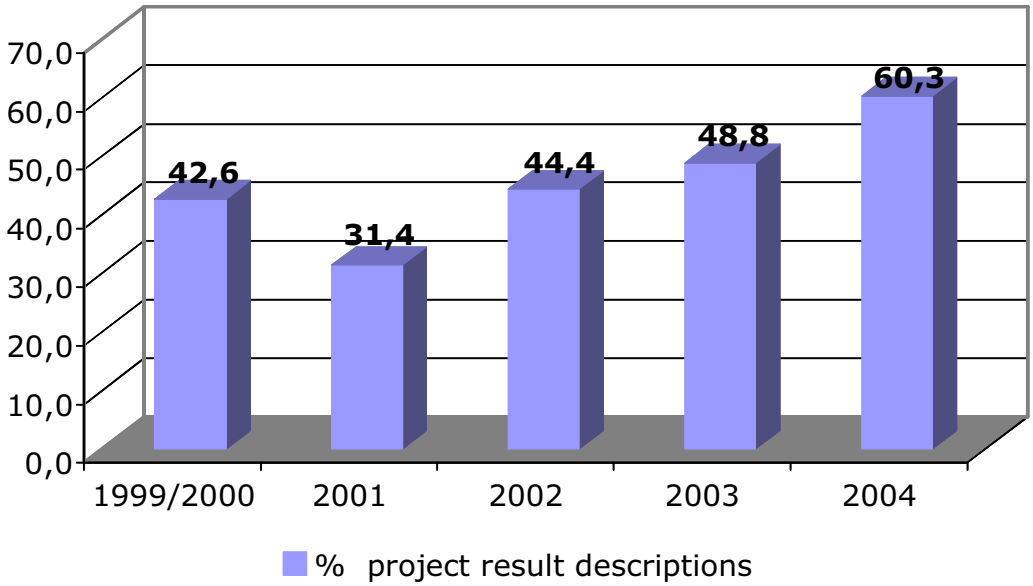


Fig. 5 - RisOSH project description in November 2007 distributed by year of project funding

Once again, more encouraging is the quality of the project descriptions that we have obtained, both as a result of our e-mails and as normal insertions by the project assignees. Looking at projects funded in 2003 and 2004 (Fig. 5), that are about to be concluded at the end of this year, we notice that there is a "natural" increase in the use of the electronic format by the project assignees: at the moment we have gathered more than 60% of the project results for the year 2004, and 48.8% of those funded in 2003. Our request to ISPESL's scientific referees has surprisingly given us more information about the results of "old" projects (42.6% for projects funded in 1999/2000), while for those funded in 2001 and 2002 the increase of information provided couldn't allow us to fill the information gap. As a matter of fact, the new mode of gathering research results through the electronic format had just been introduced in that period, and project assignees were either not used or "encouraged" by the contract to send project results electronically. The quality of the information received seems accurate: long free text descriptions of project results, both in Italian and English, descriptions of products, good use of the project classification applying ATECO and OSHA thesaurus.

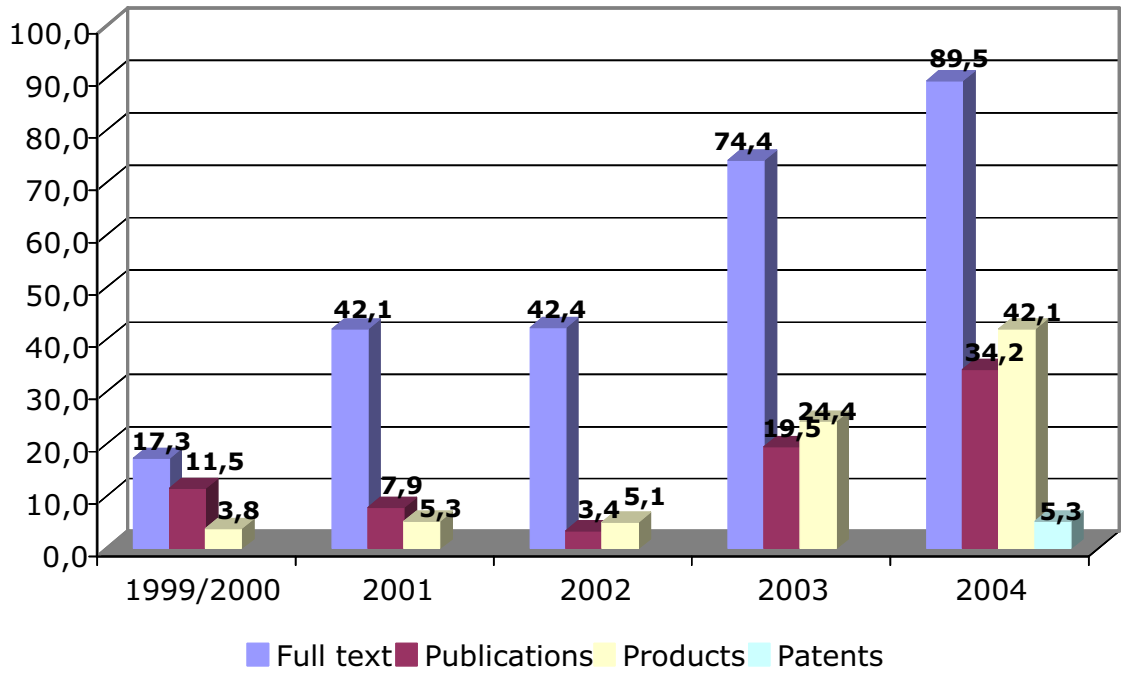


Fig. 6. – RisOSH contents in November 2007 distributed by year of project funding

Moreover, there is a high increase in the download of the full text of the final deliverables, which reach 89.5% (Fig. 6) for the projects funded in 2004 and 74.4% for those funded in 2003. This shows that researchers are getting used to share their results electronically.

Figure 6 also shows the data about publications and products. The figure reports only the presence of products and/or publications and does not give their exact number. Generally we have noticed that there is more than one publication (papers presented at international conferences and journal articles) and sometimes there are different types of products (databases and web pages, etc.), outlining a group of very "productive" and well documented project result descriptions. Interesting to note that the descriptions of products are also increasing in the last years and that they are reported more frequently than publications.

5. Discussion and conclusion

The results of our evaluation procedure show that the RisOSH system seems to be in a phase of take-off and this means that experimentations and improvements along two years can be considered very positively. The communication flow with external assignees is now working quite well and the information content of the new template has proved to be a valid tool for the collection of projects key data. In the most recent years, we can also observe that external assignees are getting more and more used to autonomously send project results.

Moreover, there is an increasing propensity to make the full-text of the final report available, that constitutes a precious source of GL documents and makes the system capable of representing an important portion of the ISPESL institutional repository. In this context, the development of the institutional communication is of outstanding importance for the realization of an effective information flow and for the planning of an editorial policy that reflects and gives added value to the scientific production of the Institute.

The improvement of such communication flow should go, however, above all toward the direction of a greater involvement of the internal community of researchers, who has shown, in our evaluation process, a good interest in the system, although not "obliged" to participate in the collection of results. As a matter of fact, the updating of the database is still performed in a centralized way because of various management reasons. What is needed is probably a greater commitment of the Institute's top management not only in "sponsoring" the RisOSH system, but also in promoting the standardization of several institutional information systems, created for different objectives, in order to make them capable of being integrated.

As a general conclusion, the evaluation carried out has stimulated our discussion on the identification of the best models to support maintenance of information systems. Actually, system updating, especially if the data entry is distributed among different information producers, constitutes one of the main problems of sustainability: suffice it to think of open access archives or institutional repositories [XIA2007, DAV2007, SURF2005]. As a matter of fact, in disciplinary fields where authors are not yet used to submit their work to open access sources, they have to be encouraged and committed to post their papers.

The open access movement, which was started on the basis of precise needs of the scholars community, and which was foreseen as a complete revolution of the traditional communication chain, first of all concerning publishers, is therefore likely to change its original organisational model, mostly in some scientific fields. In our opinion this means that between the two opposite organizational models normally found (centralised versus distributed), with the diffusion of institutional repositories on the internet, libraries and information centres are again in the front line in their role of continuous stimulation and communication with information producers, imposing a new intermediate model that requires greater flexibility of roles.

References

[ATECO2002] ISTAT, *Classificazione delle attività economiche* (ATECO 2002 der. NACE rev. 1.1.) available at: http://www.istat.it/strumenti/definizioni/ateco/note.pdf

[CER2000] Common Research Information Format CERIF 2000, *CERIF 2000 Guidelines, Final Report of the CERIF Revision Working Group Co-ordinated by DG XIII-D-D*, European Commission, http://www.cordis.lu/cerif, viewed Nov.2007.

[CRIS2006] Asserson Anne, Simons Eduard J. Eds, *Enabling interaction and quality: beyond the Hanseatic League*, 8th International Conference on Current Research Information Systems, 2006, Leuven University Press, 248 p.

[DAV2007] Davis Philip M., Connolly Matthew J.L., *Institutional Repositories. Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace*, 2007, D-Lib Magazine, 13 (3-4).

[EUR2004] EuroCRIS 2004 Conference, *Putting the Sparkle in the Knowledge Society*, Antwerp, 13-15 May, 2004, http://www.eurocris.org/en/taskgroups/conferences/events/cris2004/_0 .

[IAMSLIC2005] Luzi Daniela, Castriotta Maria, Manco Rosaria, Mazzucco Renza, *Grey Literature life cycle in research projects: a case study*, in: 31th Annual IAMSLIC Conference, Rome, 10-14 October 2005.

[JEF2002] Jeffery Keith G., Lopatenko Andrei, Asserson Anne, *Comparative Study of Metadata for Scientific Information:The place of CERIF and Scientific Repositories*, Proc. CRIS 2002 Gaining Insight from Research Information, Aug.29-31, Kassel, http://www.uni.kassel.de/CRIS2002/programme.ghk .

[LUZ2004] Luzi D., Castriotta M, Manco M.R., Mazzucco R. *The integration of GL documents with the research information system on occupational safety and health,* Proceedings of Sixth International Conference on Grey Literature, New York 6-7 December, 2004 (also published in "Publishing Research Quarterly" vol. 21, nr. 1, 2005).

[SURF2005] van Westrienen Gerard, *Completed Questionnaires, Country Update on academic Institutional Repositories*, CNI-JISC-SURF Conference, Amsterdam, 10-11 May 2005.

[VAN2004] Van de Sompel Hebert, Payette Sandy, Erickson John, Logoze Carl, Warner Simeon, *Rethinking Scholarly Communication. Building the System that Scholars deserve*, D-Lib Magazine, vol. 10, no. 9, Sept. 2004, http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html.

[XIA2007] Xia Jingfeng, Sun Li, *Assessment of Self-Archiving in Institutional Repositories: Depositorship and Full-Text Availability*, 2007, Serials Review, 33(1):14-21.

Open access to full text and ETDs in Europe: improving accessibility through the choice of language?

Christiane Stock
INIST-CNRS, France

Abstract
Next to journal articles and eprints electronic theses and dissertations (ETDs) are the most frequent document type found in open archives, for various reasons:
- ETDs are a well defined and well referenced document type. Rules for deposit and citation are generally established on a national level, and international standards exist for specific information and theses metadata, contrary to other grey documents.
- ETDs are administrative documents, and students can be "obliged" to deposit their work in an archive or repository for formal reasons.
The paper describes some tendencies concerning electronic theses and dissertations in Europe as observed during explorations of institutional and other repositories, with specific regard to the full text.
In the first part we examine the changing landscape in repositories where access to the complete full text of a thesis no longer is the unique offer, but where we also find partial access, temporary embargoes and bibliographic citations only. The increase of the number of items thus appears somewhat counter-balanced by a decrease in "quality".
In the second part, we take a closer look at the language issue of ETDs.
If the technical open access to the full text of ETDs is increasing, it doesn't necessarily imply an easier access (i.e. readability) for everyone, since the language barrier may still exist. Can the "accessibility" be increased by choosing the right language?
Today we observe a growing number of ETDs written in English, the vernacular language for scientific research. But who writes in English? In this paper we explore the phenomenon under different aspects.
- Are there differences between countries?
- Are there differences between disciplines?
- Are foreign students more inclined to write in English?
- Does the participation in international projects influence the choice of language?
The growing complexity of the ETD landscape calls for explicit policies to inform the user of a given repository as well as for tools such as directories providing detailed information on a general level.

FROM OPEN ARTICLE ARCHIVES TO COMPLEX REPOSITORIES

In the early days of the open archive movement the following equation was true for many sites: Archive = Journal Article = Open Access to its full text = Self-deposit by the author

The archive/repository might be disciplinary, like arXiv, or institutional. The article might be available in its preprint or its postprint version, but free access to the entire full text was guaranteed. Since the author submitted his article, legal obstacles only concerned the issues with the commercial editor.

The landscape has changed very much since these early days. Soon other document types were added, in particular doctoral theses and conference papers. Institutions set up policies to give open access to the entire scientific production. Today the Directory of Open Access Repositories (OpenDOAR) uses the following categories for repositories: aggregating, disciplinary, governmental, institutional (OpenDOAR 2007: http://www.opendoar.org/find.php). Each category can be combined with document types and/or disciplinary criteria.

Technical tools like e-prints and DSpace software platforms are increasingly used to store (and archive) the scientific output of a given community, whether it is in public access or not. The presentation of Andrew (Andrew 2006) shows examples of the growing complexity of repositories and portals to electronic resources. He classifies repositories as follows:

Figure 1: Basic repository classification by Andrews

Moreover, scientific publications, including grey literature, are no longer the only items to populate repositories. An increasing number of websites add research material in text form or multimedia

recordings, mixing publications and archives. One example is the University of Cambridge Institutional Repository <http://www.dspace.cam.ac.uk/>, which stores digital information in its widest sense. The same observation applies for the Loughborough University's Institutional Repository <https://dspace.lboro.ac.uk/dspace/>. The aforementioned platforms have been essential to increase and improve access to scientific material in any form by providing user friendly interfaces and supporting non-textual material.

ETDS IN EUROPEAN REPOSITORIES:

ETDs have been the second type of scientific publications to make its way to open archives or repositories, for various reasons (Paillassard 2005).

- ETDs are a well defined and well referenced document type. Rules for deposit and citation are generally established on a national level, and international standards exist for specific information and theses metadata, contrary to other grey documents.
- ETDs are administrative documents, and students can be "obliged" to deposit their work in an archive or repository for formal reasons.

The following table gives an overview of repositories registered with Open DOAR and containing theses. OpenDOAR was chosen because information provided during registration is systematically reviewed by the administrators. We find a consistent use of the different categories and document types throughout the entries which provides a good basis for comparisons.

Country	Repositories registered	Repositories with theses	%
Austria	6	3	50
Belgium	17	15	88
Croatia	3	2	66
Czech Republic	2	0	0
Denmark	6	4	66
Estonia	2	1	50
Finland	8	6	75
France	34	18	53
Germany	117	88	75
Greece	5	3	60
Hungary	4	0	0
Iceland	2	1	50
Ireland	6	3	50
Italy	28	18	64
Netherlands	44	16	36
Norway	6	5	83
Poland	9	2	22
Portugal	5	4	80
Russian Federation	2	0	0
Serbia and Montenegro	1	0	0
Slovenia	1	1	100
Spain	18	11	61
Sweden	31	31	100
Switzerland	6	4	66
Ukraine	5	2	40
United Kingdom	108	34	32
Europe	**476**	**272**	**57**

Table 1: Source: OpenDOAR (data collected October 31st, 2007)

According to the OpenDOAR registry 272 out of 476 repositories (or 57%) in 26 European countries contain theses. Only in 6 countries the part of repositories with theses is inferior to 50%. At this stage no distinction is possible between master's theses and doctoral theses.

It should be noted that not all repositories with ETDs are registered in OpenDOAR. On the other hand, some sites included in the table may still be at an experimental stage and contain either very few records or only test records. To a certain degree the description of a given repository in OpenDOAR provides valuable information on what to expect from the website.

ETDS AND QUALITY OF ACCESS TO THE FULL TEXT

In the following part we examine the quality of the information available in open access.
Indeed the equation "archive/repository = 100% full text" no longer works, instead the landscape has become rather complicated.

Many academic and research institutions have defined or are defining policies, declaring the deposit of scientific works in their institutional (or national) repository a "passage oblige" or mandatory step for either the evaluation of a researcher, or the defence of a doctoral dissertation. It is the case for master and PhD students - from 2005 onwards - at the University of Edinburgh (www.era.lib.ed.ac.uk) Similar academic regulations are found at many other universities like the University of Leiden (Netherlands) (www.promoveren.leidenuniv.nl/index.php3?m=41&c=48) or the University of Liège (Belgium) starting in 2006 (http://bictel.ulg.ac.be/presentation_2.html). The need for restricted access to parts of the documents may create obstacles to these policies, but different kinds of solutions have been developed. Some examples will be shown in the following part.

a) The full text exists, but access is restricted

If a deposit in the institutional repository is the "passage oblige" before or after the defence of a thesis, the authorization for worldwide dissemination of the full text does not necessarily follow. Aspects like confidentiality or quality criteria may lead to a restricted access to the document itself while the metadata are available in most cases. Users of such repositories find the information at different levels or stages of their visit.

- The repository of Liège University on Bictel/e, a Belgian portal for theses from 9 French-speaking universities, informs the user through a list per faculty of thesis in public or restricted access.

Parcourir toutes les thèses de BICTEL/e-ULg par Departement

Département	Type d'accès			Total
	Public	Restreint	Mixte	
Droit - Département de droit	1	2		3

Figure 2: Liège University
<http://bictel.ulg.ac.be/ETD-db/collection/browse/by_departement/>

- The Catholic University of Leuven (KUL - Katholieke Universiteit Leuven) provides a general information on the possibility of restricted access on the homepage <https://repository.libis.kuleuven.be/dspace/libridoc/index_en.jsp> and puts an information at the end of the record telling that the document is only available through the intranet.

Type	Doctoral thesis Doctoraatsthesis
Accessibility	Only internally accessible from the K.U.Leuven intranet
Toegankelijkheid	Enkel intern toegankelijk via het K.U.Leuven-intranet

Figure 3: Record from KU Leuven (part) – access restricted to the intranet

b) Partial access to the full text

Authors may be charged with the submission process only once. What to do if the thesis is declared confidential in part or in its entirety? What procedure to adopt if the dissertation is based on articles for which restrictions exist or the moving wall applies?

On the technical level software platforms allow the author to deposit his document once, but separated into different files, and to declare some parts under embargo. A user from outside the campus may come across a doctoral thesis for which the table of contents, introduction and conclusion as well as part 1 and 2 are freely accessible, but the link given for the other parts won't work for him. The University of Leiden has adopted this practise for its repository. Figure 4 shows how the end user is informed about the restrictions. Even the final date for the parts under embargo is mentioned.

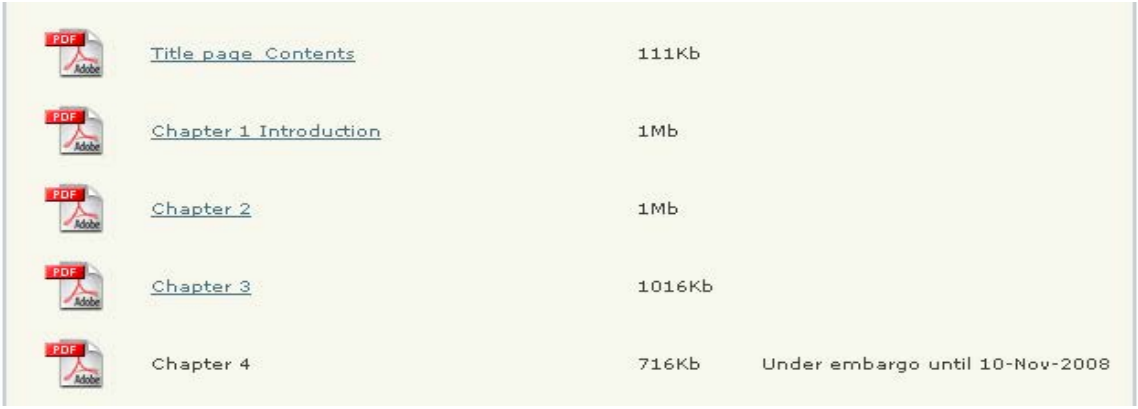


Figure 4: Leiden Repository : parts under embargo
<https://openaccess.leidenuniv.nl/handle/1887/4460>

Finland offers a different solution. Many Finnish Doctoral theses are based on articles. The full text documents accessible through the repository list the articles in the table of contents and include a summary in the corresponding chapter. The publications seem to be joined as appendices to the thesis, but are not included in the open access version.

Another way to deal with restricted parts of ETDs is shown by the University of Oslo repository DUO. The bibliographic record includes the list of papers with a hyperlink to the commercial publisher. So the end-user may access to these parts if he has a subscription for the journal or if he's willing to pay for it (example : <http://wo.uio.no/as/WebObjects/theses.woa/wo/0.3.9>)

c) Records without theses

To support authors with their deposits and to increase their willingness to participate, many repository administrators have started to upload citations from external sources (SCI, Pubmed, library catalogues, etc.). The only task left for the author is to validate the information and to upload the full text. This turn off from the self-deposit practises creates new problems and inconsistencies in the repositories like duplicate records or metadata entries without any full text.

Indeed an increasing number of repositories may give access to bibliographic records only, some of them for a limited period of time. In some cases the metadata for ETDs originate from library catalogues. We came across some cases with duplicate entries – metadata added in the deposit process and uploaded from a bibliographic file in the Ghent University archive (e.g. Van Daele, I: Identification of genes related to self-incompatibility in ryegrass (Lolium perenne L.): <https://archive.ugent.be/handle/1854/5391> and <https://archive.ugent.be/handle/1854/3915>)

While some institutions store the entirety of records on the platform and provide public access only for those documents which are complete with the full text and not under embargo, other sites go public with all their records, with the aforementioned discrepancies. As long as these differences are clearly indicated, they may be acceptable for the end-user. Often enough though he discovers the missing text at the last minute, when he displays the bibliographic record. Again, institutions use different ways to inform the public about the missing document:

At the start of the search :
- A pull down menu in "adjusted browsing" at the University of Tampere (Finland) offers the possibility to exclude records with medadata only <http://acta.uta.fi/english/kenttaselaus.phtml>.
- A check box requesting "only full text results" in the basic research (absent for the advanced search) on the research platform Alexandria of the University of St Gallen (Switzerland) <www.alexandria.unisg.ch/publications/basic-search/L-en>. DUO at Oslo University offers a similar possibility for its advanced search interface <http://wo.uio.no/as/WebObjects/theses.woa/wa/advancedSearch?>.

At the display of the list of results:
- Addition of a symbol in the citation (Hasselt: <https://doclib.uhasselt.be/dspace/>)
- A column indicating the file format to inform that the full text is available at Uppsala University

d) Other solutions:

The Imperial College London Repository has chosen to give public access only to validated full-text dissertations, while an in-house platform is used for the workflow and administrative purposes (Jones 2007).

In France all doctoral theses must be referenced in the national bibliography, whereas access to the full text (paper or electronic form) may be "confidential". The present metadata scheme TEF (TEF 2007) provides data for two versions of the doctoral thesis: the official one which is confidential, and a reduced version stripped of its confidential parts for dissemination.

What happens if the repositories are harvested by service providers?
Several checks with OAIster showed that in some cases the University of Michigan service succeeds in eliminating OAI records without full text (e.g. Lund), in other cases not. Hasselt University exposes to metadata harvesting only those documents for which the full text is available (Goovaerts 2007).

We have seen that while technical developments have greatly facilitated the dissemination of ETDs, increasing their visibility to a worldwide level, legal issues, especially copyrighted parts of third authors used in the thesis, became an obstacle. Different solutions have been found and adopted so far.
The French metadata scheme TEF allows to identify and exclude passages not to be disseminated, or anticipates the co-existence of a complete version and a public version. Other Universities use the embargo scheme for parts of the document. The worst option occurs when no full text is available at all and the user learns about the absence at the last minute.
From a user's perspective it would be very helpful if repositories stated clearly their policies on what is freely accessible, for each collection/discipline as well as in the "About" text of their site and in their descriptions for registries such as OpenDOAR and ROAR.

ETDS AND LANGUAGE ISSUES :

When looking for ETDs written in English for the thematic portals INIST proposes for its research communities, we observed that their number seemed to be increasing in the German repositories. Curiosity arose about whether this was a generalized tendency throughout Europe. Indeed, what good is a worldwide access to a document if you don't understand the language and can't read the text? On the other hand, writing your thesis in English (or French) may give openings to a wider range of readers.

a) Methodology, limits and difficulties

The following methodology was used for our survey: we selected repositories referenced either in the OpenDOAR or the ROAR (Registry of Open Access Repositories – <http://roar.eprints.org/index.php>. A quick check allowed us to find out if the following elements could be identified and isolated: document type = theses or dissertation, year of publication, the language. Data were obtained with the help of the search and browse interfaces. Sometimes it became necessary to count the references. Whenever possible, the survey was restricted to Doctoral theses.

The study was limited to the years 2000 – 2006 (date of publication), since repositories/archives going back before 2000 are less frequent and contain only a small number of ETDs. Quite a few document servers start their activity with the years 2002 or 2003 (as date of publication). Sites like the University of Liège repository or several Italian archives had to be discarded because they were created only recently.
Our second limitation was to theses written in English as foreign language, ignoring the other languages which may occur (e.g. French, German, Italian, Spanish for some Scandinavian sites)

3 types of difficulties should be mentioned:
- Some very interesting repositories could not be included in our study, because it was impossible to either isolate the document type, or to obtain information on the language. The Catalonian portal Tesis Doctorals en Xarxa (TDX) (http://www.tesisenxarxa.net/) provides access to doctoral theses written in Catalan, Spanish, English, French etc. but it's impossible to apply language filters.
- It was difficult to obtain consistent data for the comparison between repositories. Some sites may be dedicated exclusively to PhD theses, and only to those available in electronic form in full text, while others mix PhD and master's theses, or electronic and print versions. For this part of the study online access to the full text was not relevant.
- On the repository level, we met with the following problems:
    - the possibility of duplicate entries (for various reasons),
    - the date given in the record was not the date of the defense, but rather the date of deposit or the date of digitization,
    - and the occasional check of the full text revealed some errors of the language code in the record.

The following data therefore give only general outlines.

b) A country survey

Sweden:

Scandinavian repositories include a high percentage of works written in English, but also ETDs in other European languages (French, German, Spanish etc.). We were not able to distinguish the Scandinavian languages from each other. The part of English theses is seldom below 50 %. The figure from Stockholm

University shows a phenomenon common to many repositories in other countries. For the early years when the total number of theses amounts to 10 documents or less, almost all are written in English. Currently the part of English theses turns around 70%. Lund University deposits between 360 and 470 theses per year (only part of them in full text), of which between 80 and 90 % are written in English.

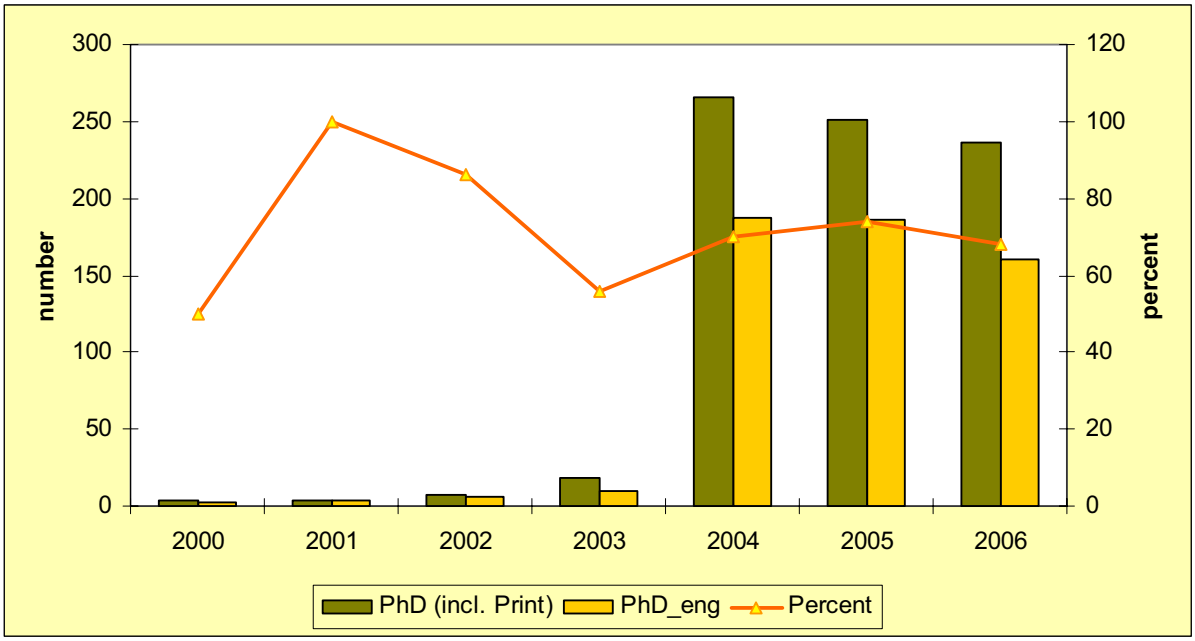


Figure 5: Stockholm University repository

Germany:
German universities had the benefit of several national programmes to set up institutional repositories and a change of academic regulations to authorize or make mandatory the deposit of electronic theses. This results in an important number of repositories as well as documents populating them. Data may vary according to the specialization of the university (biology or technical or humanities/social sciences), as well as with the size of the institution. We cumulated data from 8 universities (Bayreuth, Frankfurt, Giessen, Hamburg, Heidelberg, Potsdam, Stuttgart, and Technical University Munich). Our example shows that if the number of English ETDs is increasing, so is the overall number of documents deposited. The English part grew from 10% in 2000 to 25% in 2005, but seems to be stable now.

	2000	2001	2002	2003	2004	2005	2006
ETD Total	1259	1745	2077	2195	2376	2459	2374
ETD English	127	199	275	337	490	615	603
% in English	10	11	13	15	21	25	25

Table 2: Cumulated data from 8 German universities

France and Italy:
The methodology for the French part had to be changed. Apart from some pioneer sites like Grisemine, now succeeded by IRIS (Claerebout 2004) at Lille or Cybertheses and INSA in Lyon (Paillassard 2005) and the archive TEL (based on the author's deposit) the deposit of electronic theses in institutional repositories really started in 2005/6. We decided therefore to extract the data from the national bibliography for theses, accessible through the Sudoc catalogue <http://sudoc.abes.fr>. They include paper documents as well as electronic versions. The percentage of doctoral theses written in English can be found in the following table, but is very low compared to other countries, increasing from 0% to 3%

Sudoc	2000	2001	2002	2003	2004	2005	2006
Total	19173	16046	15770	16515	17375	17357	15122
In English	37	116	183	232	279	458	433
% in English	0	1	1	1	2	3	3

Table 3: Doctoral theses defended in French universities

In some cases the document is written in English because the thesis is defended simultaneously at a foreign university. Other checks on the site of INSA Lyon showed no obvious reasons for the choice of the English language.

We found similar disappointing results from the Italian repositories referenced in OpenDOAR: few Doctoral thesis could be identified (Firenze, Parma, Trento, Bologna). Like in France, the movement to make thesis

available through repositories is only in its beginnings (Galimberti 2007). However, all 21 theses from Trento University available online were in English.

Belgium:
2 repositories were selected for our purpose: they hold important numbers of online theses for our period: Ghent University and the Catholic University of Leuven (KUL). Both show an important part of ETDs written in English, comparable in percentage to Scandinavian repositories. Moreover the part seems to be fairly constant over the years.

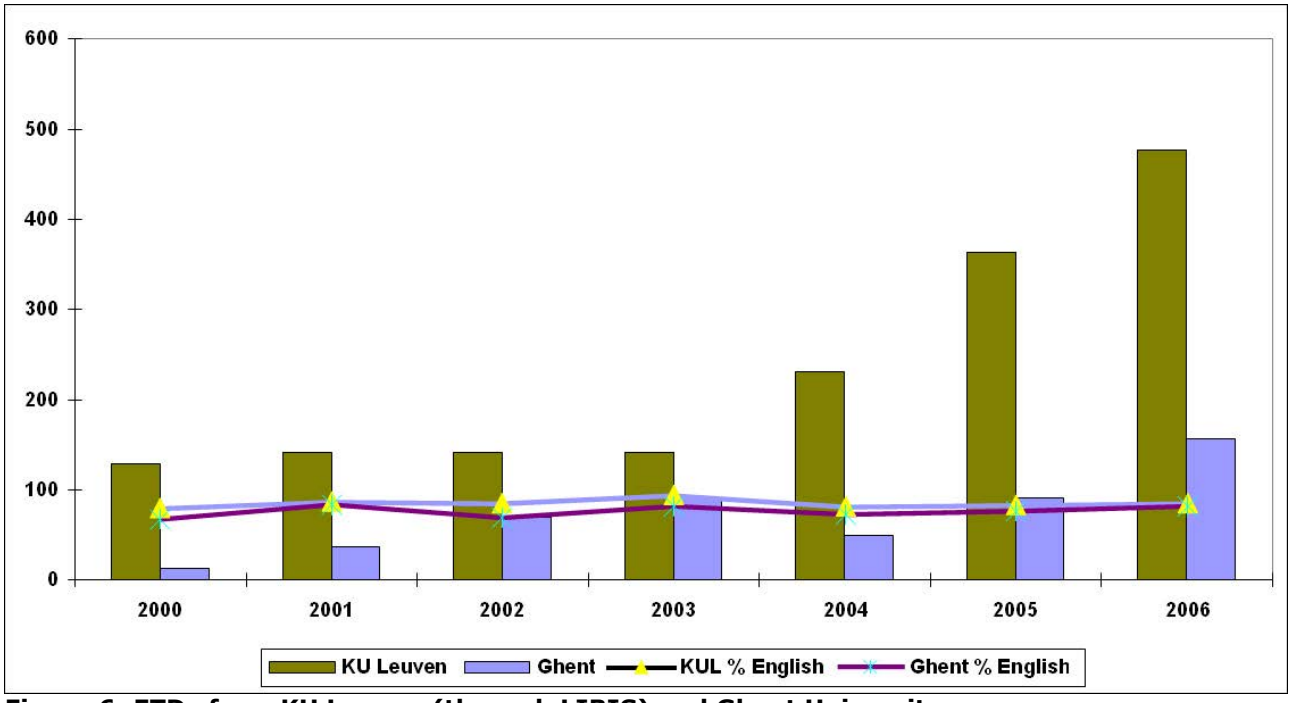


Figure 6: ETDs from KU Leuven (through LIBIS) and Ghent University

A comparison between the repositories of Lille1 University (IRIS) and Ghent University – both have a comparable number of entries - reveals the very low percentage of English ETDs in France.

Ghent (BE): 500 ETDs online - 390 in English
Lille 1 (FR): 418 ETDs online – 5 in English

A break-up by faculties at Ghent University shows the dominance of sciences, agricultural and applied biological sciences and medicine and health sciences in number of deposits and in the use of the English language. Deposits from the Faculty of engineering are also important, but the English part is much lower.
We found similar results from the rather small University of Bayreuth (Germany): apart from 1 English thesis in Cultural studies, we have 23 English documents in Mathematics and Physics and 84 in Biology, Chemistry and Earth Sciences.

In conclusion of this part we observe that there are great differences between countries with regards to the number of theses available online, as well as to the percentage of them written in English. Scandinavian countries as well as Belgium (and the Netherlands) are highly tolerant with regards to the language choice for a thesis. On the opposite site France has a low percentage of theses not written in French.

c) Who writes in English and why?

Circumstances didn't allow us to do an in-depth study of the reasons why students take up the courage to write in English. A look at the title page, the resume, introduction or acknowledgement part of the theses supplied us with some information. But this part of the study might need a closer examination. It would also take up much more time.

Some answers to the question may be obvious:
A part of the foreign students publishes in English, making thus their work accessible to the research communities in both countries. Some of the French theses written in English were submitted to 2 universities in different countries.
The difficulty of his own language (e.g. Finnish) may induce a student to write in English, thus increasing the accessibility of his work. But it doesn't explain the liberal policy Scandinavian universities have with the ETDs submitted to their universities.

Scientific domains like physics, computer science or mathematics have shown in the past their openness to share and exchange their results through the deposit in archives/repositories, and they also favour this exchange through theses written in English, followed by medical and biological disciplines. In Geosciences research conducted in a foreign country may influence the choice of language.

Last, but not least the participation of the student or his supervisor in international projects accounts for a large part of theses written in English.

CONCLUSION:

As we supposed before our study, the landscape of online access to thesis has changed in many ways in the past years.

The growing number of thesis "available" on the internet comes with an increasing diversification as to the degree level (Bachelor, Master or Doctor), the kind of access with regards to the full text and to the contents itself. Repositories mix full text entries with records without documents or theses with partial access. Institutions alert about these differences, but not in a consistent way.
In order not to disappoint the user of a given repository is it more important than ever that the administrators of the site clearly indicate the policies for the deposit of and access to electronic theses and dissertations.
Useful tools like OAIster or sites of service providers who get their records through harvesting should be examined with a critical eye if the primary need of the end-user is the access to the full text.

Our survey of theses written in English showed important differences between European repositories. In the Scandinavian countries as well as in Belgium or the Netherlands between 50 and 90 percent of (doctoral) theses are in English. German universities appear to catch up with them; as we supposed in the beginning, the percentage of English ETDs has grown to reach 25%. For several interesting repositories we were unable to get the data we were interested in, whereas others have just started to disseminate theses online, so that our survey was premature. An examination of differences based on scientific domains would be interesting, but also more difficult time consuming.
Our observation that the "pioneer deposits" of theses in repositories are mostly written in English seem to indicate the willingness to give the widest access possible to one's work: through the choice of language and through the internet.

References :

1. Andrew, Theo (2006): The Institutional Repository in the Digital Library. – Paper presented at: Institutional Repositories workshop at the DSpace User Group Meeting, Bergen, 19 April 2006. - Available at: http://dsug2006.uib.no/index.php?p=ir

2. Claerebout, M.-F. (2004), Grisemine, a digital library of grey university literature. –In: GL5 Conference Proceedings: The Fifth International Conference on Grey Literature: "Grey Matters in the World of Networked Information", 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp.27-31. - (GL-conference series, ISSN 1385-2316 ; No. 6). ISBN 90-77484-01-9

3. Galimberti, P and Marialaura Vignocchi (2007): Time for a change: the Italian CRUI-Open Access Working Group's action for a national e-theses provision service. - Paper presented at: ETD 2007: Added values to e-theses = 10th International Symposium on Electronic Theses and Dissertations, Uppsala, Sweden, 13th – 16th June 2007. – Available at: http://epc.ub.uu.se/etd2007/files/papers/abstract-22.pdf

4. Goovaerts, Marc. (October, 2007). Submission based on document type: DSpace development at Hasselt University. - 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=201&cf=11

5. Jones, R. (October, 2007). On Integration. 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=200&cf=11

6. Paillassard, P., Schöpfel, J. and C. Stock (2005) How to get a French doctoral thesis, especially when you aren't French. - In: GL6 Conference Proceedings: The Sixth International Conference on Grey Literature "Work on Grey in Progress" 6-7 December 2004. – Amsterdam :

GreyNet/TextRelease, 2005, pp. 143-151. - (GL-conference series, ISSN 1385-2316; No. 6). ISBN 90-77484-04-3

7. Sudoc: Système Universitaire de Documentation (academic union catalogue of serials and monographs): http://www.sudoc.abes.fr/

8. TEF (2007). Métadonnées des thèses françaises [Online]. - Available at : <http://www.abes.fr/abes/documents/tef/index.html>

The Repositories:

Data were collected in October and November 2007. The repositories cited can be found through the following addresses:

OpenDOAR - the Directory of Open access repositories: http://www.opendoar.org
Registry of Open Access Repositories (ROAR): http://roar.eprints.org/index.php

CORDIS
Community Research & Development
Information Service

CORDIS

Your gateway to FP7 participation

cordis.europa.eu/fp7
The 7th Framework Programme service on CORDIS

***Online services** offered by the Publications Office:*
***bookshop.europa.eu:** EU publications*
***cordis.europa.eu:** Research and development*
***eur-lex.europa.eu:** EU law*
***ted.europa.eu:** Public procurement*

SEVENTH FRAMEWORK PROGRAMME

Publications Office
Publications.europa.eu

The impact of Grey Literature in the web environment: A citation analysis using Google Scholar

Rosa Di Cesare, Daniela Luzi, Roberta Ruggieri
Consiglio Nazionale delle Ricerche, Istituto di Ricerche sulla Popolazione e le Politiche Sociali

1. Introduction

The use of Grey Literature (GL) has hitherto been studied on the basis of whether and to what extent GL documents had been cited in peer-reviewed conventional literature applying citation analysis techniques based on primary sources (i.e. bibliographic references of journal articles) [Alb2000, Dic2004], or on the multisciplinary citation indexes produced by Thomson ISI, the Web of Science (WoS) [Pel2003, Cor2004].

More recently, other tracking citation systems have been developed, such as Scopus, Google Scholar (GS), Citeseer, and CrossRef. To some extent these have challenged the ISI monopoly over scientific evaluation of countries, institutions, groups and authors. Accordingly, many studies have begun to compare the different systems, particularly WoS, Scopus and GS [Bak2006, Jac 2005, Nor2005], to reveal the citations they have in common and the discrepancies, both in terms of frequency and source. In the process, the limits of the individual systems have emerged. These include citation errors (typographical errors, non-standard reference formats, parsing errors, etc.), as well as varying focus according to discipline and the varying document types used for tracking citation counts.

Our research concentrates principally on the latter aspect and aims to ascertain the impact of GL on the web environment. Consequently we chose to use GS that considers not only citations from peer-reviewed conventional journals, but also includes citations received by GL documents. Potentially, GS is the appropriate tool for illustrating the various changes underway in scholarly communication as it can reveal not only different electronic documents, predominantly GL documents, to be found in the assortment of institutional pages, open archives and repositories, but can also reveal their impact in terms of the citations made. Therefore, a secondary aim of this paper is to verify whether GS is an efficient tool to identify core papers as well as in tracking citations from different document types and whether it is able to represent the scholarly communication deriving from citations considering both GL and conventional literature. This is fundamental for GL, which is clearly an integral part of scholarly communication for numerous scientific disciplines, but still remains on the periphery of research evaluation activities.

The paper presents the results of a citation analysis with the aim of evaluating the use of GL through a comparison of the number of citations received by both GL and conventional documents. The analysis is carried out retrieving documents dealing with the topic *population ageing,* a "hot multidisciplinary topic", which is studied under demographic, socio-economical or medical perspectives.

2. Methodology

We have chosen to carry out the citation analysis using GS because this tracking citation index is:

- The only open source system, freely available on the Internet,
- It claims to have a multidisciplinary subject coverage especially in social sciences and humanities, and last but not least
- It tracks citations to GL.

Ageing is particularly suited to citation analysis as it has been a hot topic in recent years. Research on ageing has been carries out in our Institute, where authoritative bibliographic material is available. Moreover, many governmental and intergovernmental institutions carry out research on this topic under various perspectives, (demographic, socio-economic and health aspects) and they generally produce and diffuse their results using GL documents. [Dic 2006].

Our study is divided into two parts. The first one analyses the cited documents retrieved in GS in the topic chosen. We define as *cited documents* the documents that have received at least 10 citations. From the cited documents we selected a set of items, which have received at least 50 citations and we define them *highly cited documents*. In the second part we identified the citing documents of the highly cited ones focusing in particular on the relationship between cited and citing documents in terms of publication type and citation over-time.

2.1 The analysis of cited documents
To obtain our data set we used GS's search functions "phrase searching" and "allintitle" to retrieve the exact phrase "population ageing" in the title of documents. As there is also an American-English spelling, we repeated the same search using the phrase "population aging". This search strategy enabled us to retrieve fewer and more relevant results. Documents were retrieved all on the same day (13th May 2007) and we obtained 662 documents under the phase "population ageing" and 758 documents with the American-English spelling variation "population aging". From these two groups we selected the documents that received at least 10 citations. We used the GS linking function to view the cited documents in order to analyse in particular:

a) Publication type of the document: (Grey literature, Conventional literature),
b) Type of documents, (Journal article, report, Conference proceedings, etc.),
c) Date: year of publication,
d) The availability of the full-text,
e) Subject coverage.

In our sample we detected every type of GL, in particular those characteristic of the chosen disciplinary sector: working papers, discussion papers, occasional papers, which have all been grouped under the category "reports".

Regarding the topic chosen, we classified under in three categories: demo-social aspects, health aspects, and economic aspects.
The obtained data set contained two duplicates, which were excluded. The total number of documents that received at least 10 citations is 99. The results of the analysis of these documents are based on the citation counts given by GS.

From the data set of the 99 cited documents, we selected only those that have received at least 50 citations, those that we call *highly cited documents*. The resulting data set is composed of 15 documents. As one of our aims was to verify whether GS retrieves the citing documents correctly, we carefully analysed each citing document, in order to identify duplicates, or mistakes in the citation counts. The resulting data set after validation, reported in table 2, numbers a total of 885 citing documents.

2.2. Analysis of citing documents
In order to analyse the proportion of GL out of all citing items examined (i.e. frequency of GL citing) we counted the number of GL citing documents out of all the citing items analysed. We classified each of the 885 citing documents following the same criteria mentioned above (publication type, document types, etc) and further analysed the citing items considering also:

- Self-citation: depending on whether one of the authors of the citing items was also an author of the cited item,
- Language.

Particular attention has been placed on the relationship between cited and citing document types as well as their age in order to analyse the dynamics of the research field i.e. growth and obsolescence of literature and citedness in terms of immediacy and long-term impacts.

2.3. Methodological choices
GS does not have a standard format for presentation of results. Generally results presented first are more accurate in their description providing title linking to the full text or abstract, authors, source and data of publication, while toward the end of the ranking list, results are more fuzzily presented. Therefore, the retrieval of the full text was fundamental for our analysis in order to establish publication type, types of documents (especially GL documents), date of publication and verify the correctness and pertinence of citations.

To get our data set, we always used the direct link functionality provided by GS, which displays the full text, if available, or leads to a bibliographic reference generally associated with an abstract. However, since GS' data processing is done automatically, there are some incongruous descriptions, which for instance give the journal title or the report series instead of the document title, and then link to a document. In those cases we have described the retrieved document, only if it was pertinent to our data set, i.e. when the citation to the cited document was correct.

If there was no direct link to the full text or to a correct bibliographic reference, we always tried to retrieve the full text using alternatively the following GS functionalities:

- The html version;
- The "group" facility provided by GS, which "groups the multiple versions of a work [in order to] to collect all citations to all versions of work" [Goo2007] ;
- Web server address, which helped us to solve ambiguities especially on GL document types.

In this way, it was often possible to locate documents within a collection of reports produced and made available by an institution. If it was not possible to identify the type of document we followed a restrictive criteria and eliminated them from our sample.

3. Analysis of the cited documents

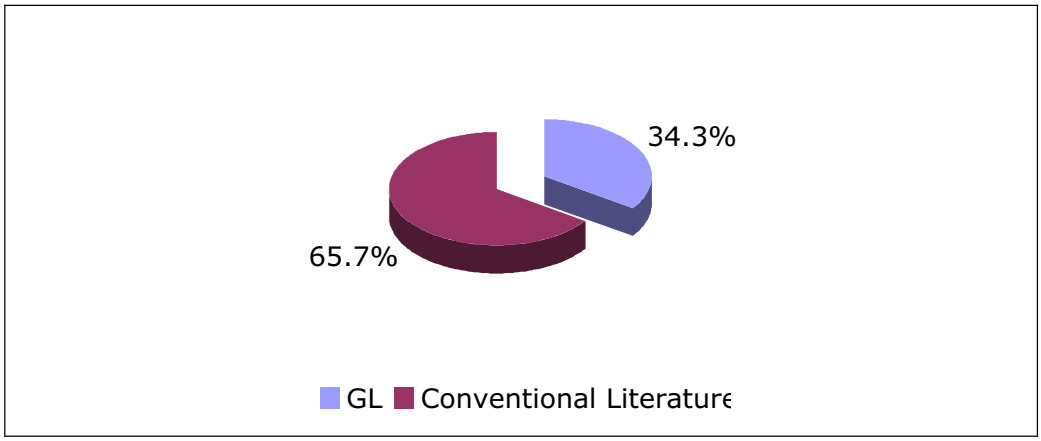


Fig. 1. – Distribution of the cited documents by publication type

Our data set is composed by 99 documents that received at least 10 citations. Figure 1 shows the distribution of the cited documents by publication type: the majority of them (65.7%, i.e. 65 documents) is published in conventional literature. In considering 34.3% of the most cited GL documents, we have to take into account that GS preferably selects the "*authoritative text* from the publisher as primary version, when multiple versions of a work are indexed" [Goo2007]. This can be misleading for the purpose of our analysis aiming to find out the role played by GL in the citation counts, as the link to the authoritative text tends to privilege conventional literature documents.

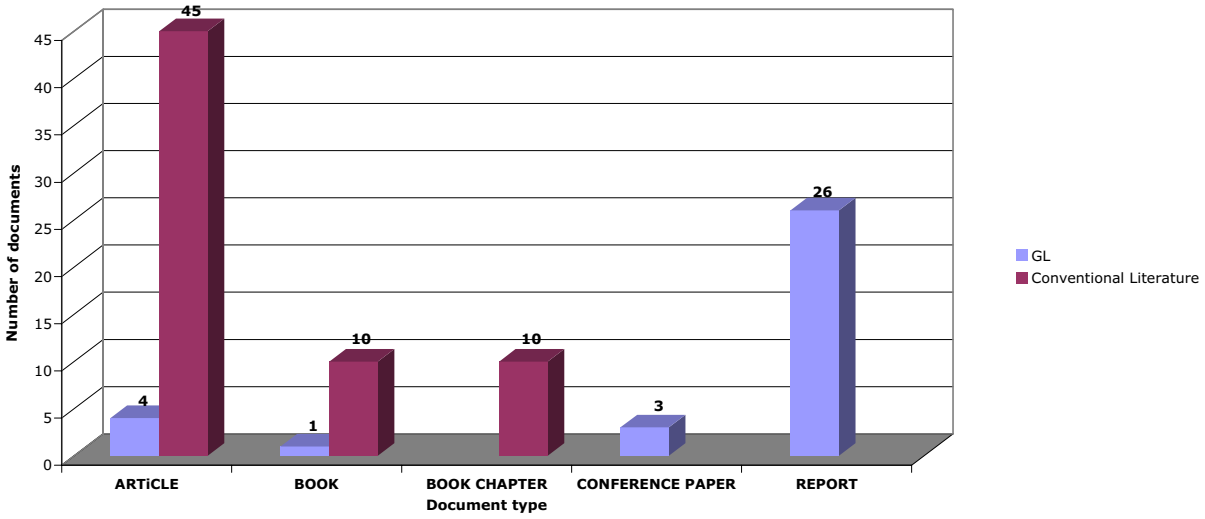


Fig. 2. – Distribution of the cited documents by document type

The distribution of the cited documents by type of document shows (fig. 2) that articles published in conventional literature are far more cited than other document types, followed by reports. The distribution by document type confirms that GS retrieves data from a variety of information sources, which include books, book chapters, conference papers and different types of GL documents, which are generally not considered in other citation tracking systems.

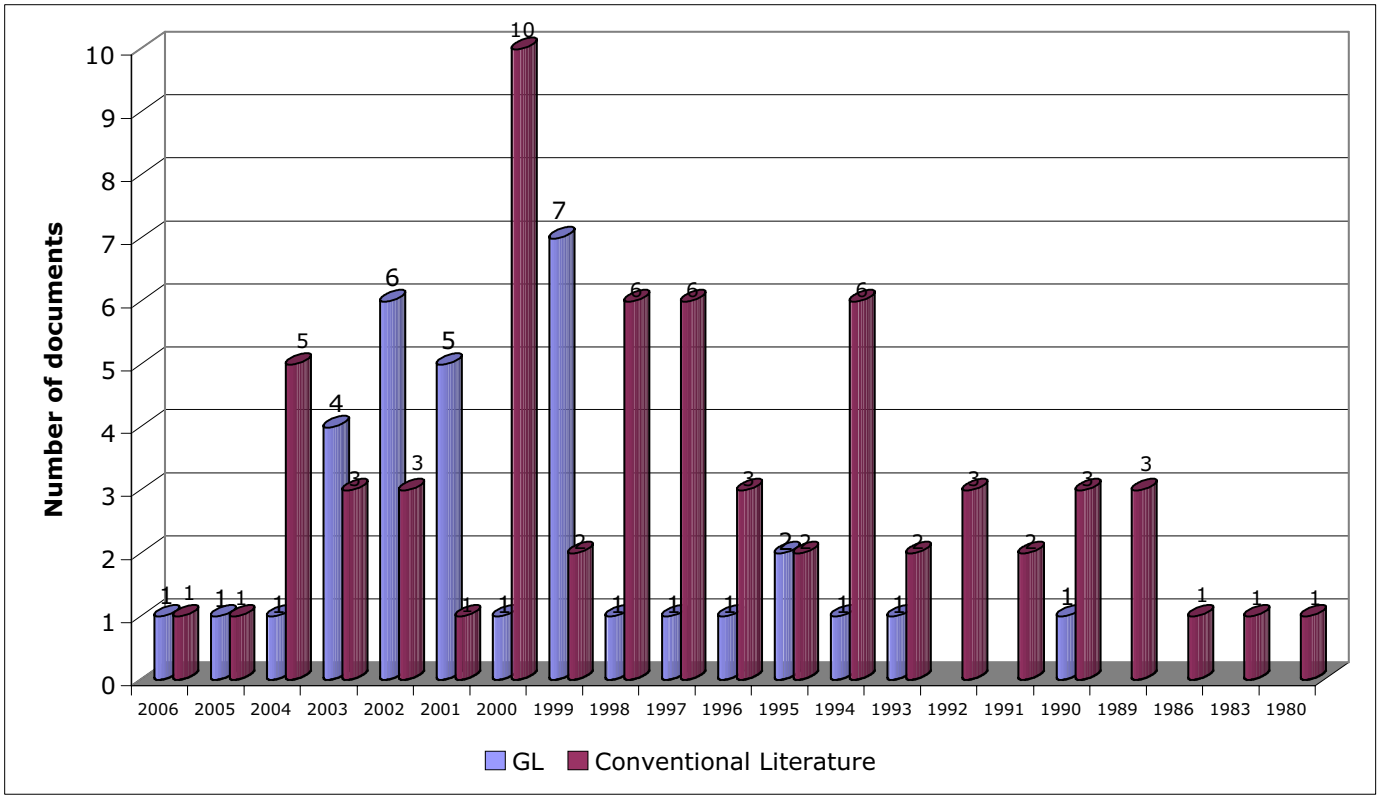


Fig. 3. - Distribution of the cited documents by publication year and publication type

Figure 3 shows the distribution of the cited documents by publication year and publication type. The GL cited documents are published from 2006 to 1993, with the exception of one document published in 1990, while the cited conventional documents are published in a larger time span, from 2006 until 1980. The highest number of conventional and GL documents were published respectively in the years 2000 and 1999. Even if GS does not declare the time coverage of its database, it can be noted that it also tracks cited documents published before Internet diffusion, the majority of them are however conventional documents.

Table 1. – Distribution of cited document by date of publication and publication type

	Total number of documents		Grey Literature		Conventional Literature	
Publication year	Frequency No.	Cumulative frequency %	Frequency No.	Cumulative frequency %	Frequency No.	Cumulative frequency %
2006	2	2.0	1	2.9	1	1.5
2005	2	4.0	1	5.9	1	3.1
2004	6	10.1	1	8.8	5	10.8
2003	7	17.2	4	20.6	3	15.4
2002	9	26.3	6	38.2	3	20.0
2001	6	32.3	5	52.9	1	21.5
2000	11	43.4	1	55.9	10	36.9
1999	9	52.5	7	76.5	2	40.0
1998	7	59.6	1	79.4	6	49.2
1997	7	66.7	1	82.4	6	58.5
1996	4	70.7	1	85.3	3	63.1
1995	4	74.7	2	91.2	2	66.2
1994	7	81.8	1	94.1	6	75.4
1993	3	84.8	1	97.1	2	78.5
1992	3	87.9		--	3	83.1
1991	2	89.9		--	2	86.2
1990	4	93.9	1	100.0	3	90.8
1989	3	97.0		--	3	95.4
1986	1	98.0		--	1	96.9
1983	1	99.0		--	1	98.5
1980	1	100.0		--	1	100.0
Total	99	--	34	--	65	--

50% of cited documents are published in a time period from 2006 till 1999. If we consider publication types, GL reaches 50% within a shorter period (2006-2001), while conventional literature is older, being published from 2006 to 1998. This means that GL cited documents are three years younger than conventional ones, indicating that GL is probably used, and therefore cited, shortly after it is produced. This is in line with one of the known GL characteristics, namely it can become obsolete in a short time or it is transformed in conventional literature.

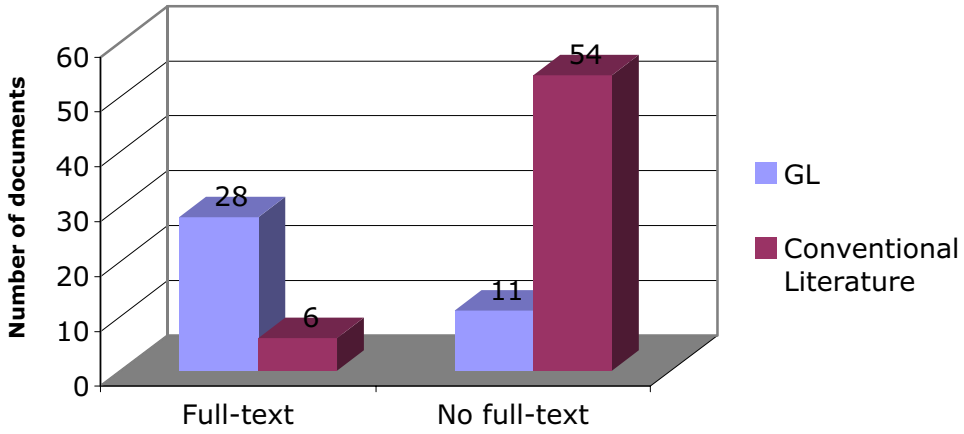


Fig. 4. - Distribution of the cited documents by publication type and full-text availability

Data on full text distribution by publication type is not surprising: (fig. 4): the majority of GL documents is available full text, whereas for conventional literature, the opposite is true. It is perhaps more interesting to analyse GL documents for which the full text is not available, along with conventional literature documents for which it is. Availability in electronic format would seem to be independent of publication date. Among the GL documents, only one was published in 1990, a period when Internet use was not widespread, whereas the conventional literature documents were published between 1997 and 2004, with the exception of one from 1989. The latter category includes numerous documents from periodicals, including the well-known British Medical Journal, Health affair, which adopt the policy of making certain articles freely available.
Finally, our data set cannot be said to reveal a correlation between full text availability and the number of citations received by documents as certain researchers have claimed [Hit2002].

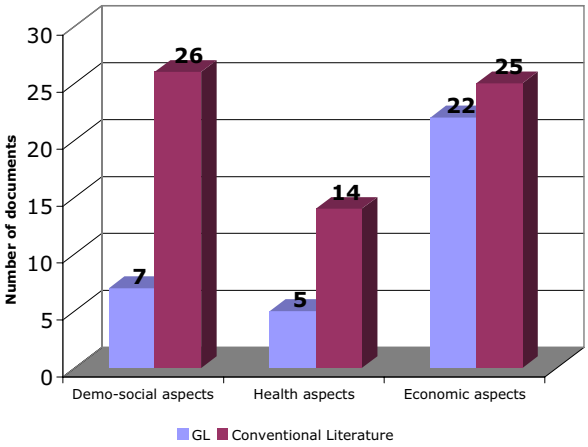


Fig. 5. – Distribution of the cited documents by topic and type of publication

The topic population ageing can be studied under various perspectives, which have been pointed out in our classification. Population ageing focusing on demo-social aspects gathers documents, which deal with major issues of population studies: fertility, mortality, immigration, family structure and social policies and generally contain statistical data. Health aspects concern the impact of population ageing on the organisation and costs of health care systems as well as particular pathologies connected with the elderly, while Economic aspects give particular attention on issues related to the labour market, pension and fiscal systems. In our sample the majority of cited items focus on economic aspects and this is also the topic where GL (22 documents) and conventional literature (25) have a similar amount of documents. The function of GS in promoting visibility of open access material is particularly evident in this figure, and even

if it is not possible to verify GS comprehensiveness, its capability of tracking documents in institutional web sites and/or open archives can surely help to diffuse the production of economic reports to a wider audience.

4. Profile of the highly cited documents

The following paragraphs deal with the analysis of the 15 highly cited documents related to their corresponding citing items. As mentioned previously, we started the analysis with the data validation of the citing documents.

Table 2. – Data validation of documents citing the 15 highly cited documents

Documents	Type of document cited	GS citing documents	Duplicates		Missing reference		Incomplete bibliographic elements		Citations analysed
		No.	No.	%	No.	%	No.	%	No.
P1	ARTICLE	158	6	3.8	--	--	5	3.2	147
P2	REPORT	105	11	10.5	2	1.9	6	5.7	86
P3	REPORT	101	7	6.9	3	3.0	8	7.9	83
P4	ARTICLE	78	5	6.4	--	--	8	10.3	65
P5	ARTICLE	72	6	8.3	--	--	3	4.2	63
P6	REPORT	65	11	16.9	--	--	6	9.2	48
P7	REPORT	63	7	11.1	--	--	6	9.5	50
P8	ARTICLE	62	2	3.2	2	3.2	5	8.1	53
P9	ARTICLE	61	8	13.1	4	6.6	5	8.2	44
P10	ARTICLE	55	7	12.7	2	3.6	7	12.7	39
P11	REPORT	54	11	20.4	3	5.6	--	--	40
P12	REPORT	55	7	12.7	2	3.6	7	12.7	39
P13	ARTICLE	53	4	7.5	5	9.4	3	5.7	41
P14	REPORT	53	2	3.8	2	3.8	4	7.5	45
P15	REPORT	53	3	5.7	6	11.3	2	3.8	42
Total		**1088**	**97**	**8.9**	**31**	**2.8**	**75**	**6.9**	**885**

Highly cited documents have been selected considering only items receiving at least 50 citations. In order to check whether GS tracks citations correctly, we analysed each citing document retrieving their full-text. Table 2 shows the results of data validation. In the first two columns the highly cited documents are ranked by document type, the number of citations given by GS is reported in the third column. The others show the number and percentage of documents discarded in our sample due to:

- Document duplications, i.e. documents with the same authors, title and publication year,
- Missing references, i.e. the items which did contain the bibliographic reference of the cited documents,
- Incomplete bibliographic elements, which we sorted into:
    - Documents whose bibliographic elements were not sufficient to identify whether it belonged to GL or conventional literature, and/or the document type of the citing item;
    - Documents written in oriental languages, in which we could not check the correctness of citation and/or identify document type or any other important information necessary to classify the citing documents.

The last column shows the number of citations received by the highly cited documents after data validation, on which our analysis is based.

The majority of discarded documents are duplicates (97 documents, 8.9%): GS automatically tracks documents available in different web pages without making any double counting citation control. We carefully analysed these documents and included in our sample only those that had remarkable differences, for instance the same title and different authors and/or different date of publication, i.e. we considered the different versions of the same document. GS proudly claims that it can trace all the various versions of a document enabling a user to retrieve the complete text. This function can also throw up important information on a document's life cycle, thereby often consenting its transformation from GL to conventional literature. It is certainly true that for citation analysis this entails a time consuming activity in order to obtain a coherent sample of data. Incomplete bibliographic elements were found in 6.9% of citing documents, where the retrieval of the full text, especially in GL documents, did not enable us to establish publication and types of citing documents.

A limited number of citing documents were discarded because they did not contain the citations to our selected documents (31 documents, 2.8%). In the end 81.3% of GS citing items were valid, while 18.7% were discarded. Note that considering all the discarded documents, a larger number of citations to GL documents were missed and this depends on the usual weak points in the bibliographic description of GL documents.

Table 3. – Profile of the highly cited documents

Highly cited documents	Reports		Journal articles		Total	
Citation rate	No.	%	No.	%	No.	%
≥80	2	13.3	1	6.7	3	20.0
50-79	1	6.7	3	20.0	4	26.7
< 50	5	33.3	3	20.0	8	53.3
Total	8	53.3	7	46.7	15	100.0
Date of publication						
1990-1994	2	13.3	3	20.0	5	33.3
1995-1999	3	20.0	2	13.3	5	33.3
2000-2003	3	20.0	2	13.3	5	33.3
Impact						
Cited in the same year of publication	5	33.3	2	13.3	7	46.7
Cited in other years	3	20.0	5	33.3	8	53.3
Cited in 2007	5	33.3	5	33.3	10	66.7
Cited in other years	3	20.0	2	13.3	5	33.3
Full text						
Available	6	40.0	2	13.3	8	53.3
Not available	2	13.3	5	33.3	7	46.7
Topic						
Health aspects	2	13.3	3	20.0	5	33.3
Economic aspects	5	33.3	2	13.3	7	46.7
Demo-social aspects	1	6.7	2	13.3	3	20.0

Out of 15, 8 are GL documents, most of which are produced by international institutions (OECD, UN, IMF), while conventional journal articles are published in Journals indexed by ISI (Table 3). Grey documents are among the top ranking items that receive the highest number of citation (more than 80 citations), however, the majority of reports receive less than 49 citations. Citations received by conventional journal articles are more homogenously distributed.

No remarkable difference exists in the date of publication of our sample, while there are some differences in their citation impact if considering citations received in the same year of publication. A more detailed analysis on this topic can be found in table 8 and in figures 7 and 8.
GL is not always available in full text, nor do conventional documents always have a restricted access, as we might have expected.

In our sample there is a slight majority of documents covering economic aspects, most of them are reports, while a relevant part of conventional journal articles covers health aspects (33%).

5. Profile of citing documents

In this paragraph we present a short profile of the citing items based on the variables selected for our analysis, which can help us in finding out the correlation between cited and citing documents.

Table 4. – Profile of citing documents

Characteristics of citing items	All items citing		Publication type: Grey literature		Publication type: Conventional literature	
	No.	%	No.	%	No.	%
	885	**Total**	**523**	**59.1**	**362**	**40.9**
Documents type	No.	%	No.	%	No.	%
Journal article	416	47.0	72	13.8	344	95.0
Report	302	34.1	302	57.7	--	--
Conference paper	85	9.6	85	16.3	5	1.4
Thesis	33	3.7	33	6.3	--	--
Draft	21	2.4	21	4.0	--	--
Book and book chapter	28	3.2	10	1.9	13	3.6
Age	No.	%	No.	%	No.	%
Less than 5 years	544	61.5	310	59.3	234	64.6
5 years - 10 years	285	32.2	186	35.6	116	32.0
More than ten years	31	3.5	8	1.5	6	1.7
Not available	25	2.8	19	3.6	6	1.7
Citation rate	No.	%	No.	%	No.	%
Self-citation	60	6.8	33	6.3	27	7.5
Citation by other	825	93.2	490	93.7	335	92.5
Language	No.	%	No.	%	No.	%
English	762	86.1	426	81.5	336	92.8
# English	123	13.9	97	18.5	26	7.2
Full text	No.	%	No.	%	No.	%
Available	577	65.2	499	95.4	78	21.5
Not available	308	34.8	24	4.6	284	78.5
Subject	No.	%	No.	%	No.	%
Health aspects	367	41.5	154	29.4	213	58.8
Economic aspects	362	40.9	272	52.0	90	24.9
Demo-social aspects	156	17.6	97	18.5	59	16.3

885 documents cite 15 highly cited documents (table 4), and most of the citing items are GL (59.1%, 523 documents), while the remaining 40.9% belongs to conventional documents (362 documents). In GL citing items, reports are the most citing type of documents, while conventional journal articles represent almost entirely the citing sample within conventional literature. More than 60% of the citing items have been published in the last six years and 32.2% is no older than ten years. Our sample is only marginally influenced by the self-citation phenomenon (6.8%).

The citing documents are predominantly written in English and languages other than English are prevalently GL citing documents (18.5%). The majority of the citing items (65.2%) links to full-text, as expected most of them are GL (95.4%). Regarding the subject coverage, most of the citing documents cover both economic and health aspects. This figure is comparable with data on subject coverage of the highly cited documents (see tab. 3).

Table 5. - Distribution of citing documents by date of publication and publication type

	Total		Grey literature		Conventional literature	
	Frequenc No.	Cumulativ frequency %	Frequenc No.	Cumulativ frequency %	Frequenc No.	Cumulativ frequency %
2007	26	3.0	9	1.8	17	4.8
2006	137	19.0	81	17.9	56	20.5
2005	127	33.7	85	34.7	42	32.3
2004	136	49.5	78	50.2	58	48.6
2003	118	63.3	57	61.5	61	65.7
2002	108	75.8	75	76.4	33	75.0
2001	71	84.1	34	83.1	37	85.4
2000	54	90.3	32	89.5	22	91.6
1999	27	96.6	20	93.5	7	93.5
1998	25	96.4	16	96.6	9	96.1
1997	17	98.4	9	98.4	8	98.3
1996	3	98.7	2	98.8	1	98.6
1995	5	99.3	3	99.4	2	99.2
1994	2	99.5	1	99.6	1	99.4
1993	1	99.7	--	--	1	99.7
1991	3	100.0	2	100.0	1	100.0
Total	860		504		356	
Not availabl	25	2.8	19	3.6	6	1.7
Total	**885**		**523**		**362**	

The cumulative frequency of citing documents (Tab. 5) makes it possible more precisely to outline the age of citing documents. We found that more then 50% of the citations given to the most cited papers were concentrated in the last five years (63.3%), with the maximum value in 2006 (137 citing documents). It is interesting to note that a large proportion of citing items were no older than four years (49.5%). On this point, the tendency of demographic sciences to feature a higher concentration of scientific literature in the recent past has already been highlighted in a previous survey [Dic2006].
The general trends of the GL and conventional literature are not very different. La LG non è come ci si aspetterebbe più giovane della Letteratura convenzionale (se non una esigua differenza nei valori percentuali del 2005 e del 2004). Contradicting what one might presuppose, GL is not younger than conventional literature (with the exception of slight differences in the percentages for 2005 and 2004). Finally only 2.8 of the citing documents lacked the date of publication. This percentage is higher for GL citing documents (3.6%), which goes to confirm the difficulties inherent in the bibiliographic control of this type of document.

6. Correlation between cited and citing documents

In the analysis of the correlation between cited and citing documents we have focused our attention on two important variables, namely the publication and document types and date of publication.

The former intends to verify whether and to what extent publication and document types influence the citation practice and in particular if a specific document type tends to cite the same type of document. For instance Cronin et al. [Cro1997] verified the "tendency of books to cite books and articles to cite articles". It would be interesting to carry out long term monitoring of the phenomenon to see whether greater on line availability of a wider range of document types (book, book chapters, conference proceedings and reports) could modify citation behaviour including the varying output of scientific activity in the scholarly communication.

The latter variable is a classical element of citation analysis. The link between publication dates of the cited and citing documents gives precious insight on the more or less immediate impact of a publication and, more broadly speaking, on the factors influencing its citedeness and life cycle (latency, immediateness, long term impact etc.).

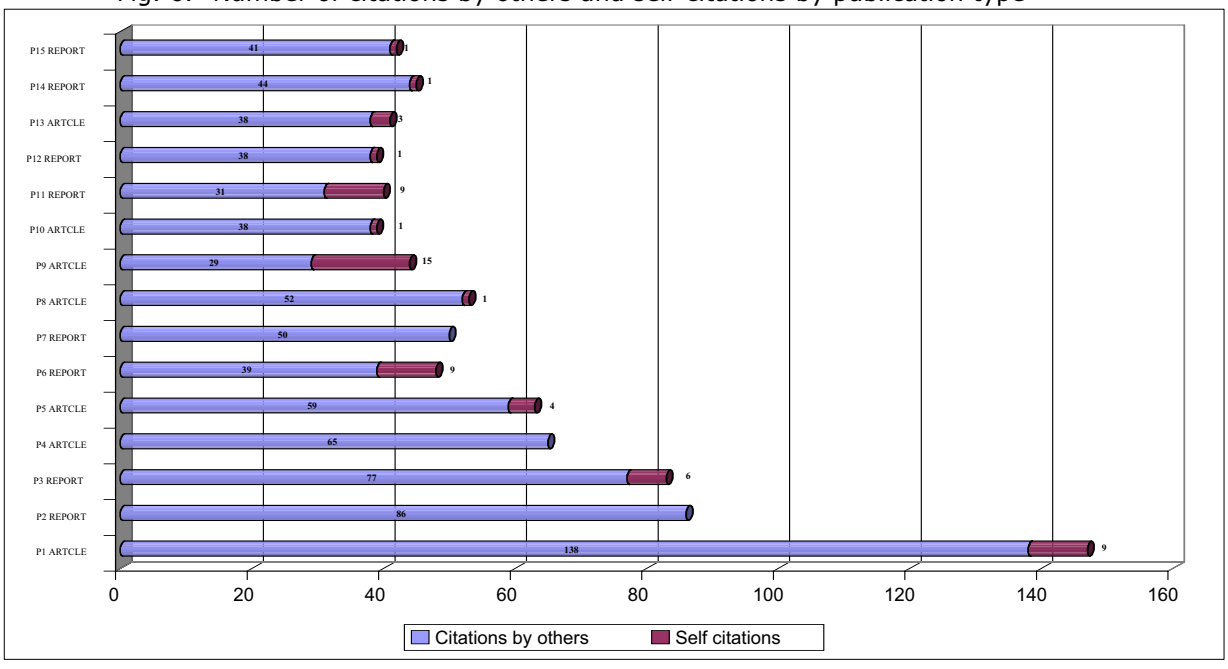
6.1. Citation links of document types

Table 5- Distribution of citing documents by publication type

Documents	Type of document cited	Total number of citations	Grey literature		Conventional literature	
			No.	%	No.	%
P1	ARTICLE	147	81	55.1	66	44.9
P2	REPORT	86	54	62.8	32	37.2
P3	REPORT	83	61	73.5	22	26.5
P4	ARTICLE	65	14	21.5	51	78.5
P5	ARTICLE	63	18	28.6	45	71.4
P6	REPORT	48	40	83.3	8	16.7
P7	REPORT	50	37	74.0	13	26.0
P8	ARTICLE	53	28	52.8	25	47.2
P9	ARTICLE	44	35	79.5	9	20.5
P10	ARTICLE	39	6	15.4	33	84.6
P11	REPORT	40	18	45.0	22	55.0
P12	REPORT	39	30	76.9	9	23.1
P13	ARTICLE	41	34	82.9	7	17.1
P14	REPORT	45	31	68.9	14	31.1
P15	REPORT	42	36	85.7	6	14.3
Total		**885**	**523**	**59.1**	**362**	**40.9**

The Table 5 shows the distribution of the citing items and the number of citations by type of document cited. In the first two columns the 15 highly cited documents are ranked by Article or Report. For each cited document the number and distribution of citations received are given.
Generally there is a symmetrical link between cited and citing documents: a GL cited document tends to receive citations from other GL items (see P2, P3, P6, P7, P12, P14, P15), and *vice versa* (P4, P5, P10) with percentage ranging from 60% to 80%. Two journal articles (P9 and P13) written by the same author, apply the reverse rule: they receive a higher number of citations from GL partially explainable by the high number of self-citations, which also represent the highest figure of self-citation in our sample (see fig. 6). Only in three cases did we note a more balanced relation between cited and citing items, but this does not seem to depend on the type of publication: two are journal articles (P1 and P8) and the other is a report (P11).
Although the number of our data set does not allow us to draw final conclusions, we can note that, on the one hand, in our sample GL documents in particular tend to cite more often other GL documents, while, on the other, conventional journal articles tend to cite a lower number of GL documents.

Fig. 6.- Number of citations by others and self-citations by publication type



Self- citation analysis is used to distinguish citations received by other authors from self-citations in order to identify the real impact of the paper. Broadly speaking, self-citation is less frequent in the social sciences than in natural sciences. In Snyder's 1998 survey on various disciplinary sectors, self-citation in social sciences came to 6%, reaching 15% in the natural sciences.

Our data set presents a limited number of self-citations: the percentage is 6.8% (60 documents out of 885) with no evident difference between GL and conventional literature (see Tab. 4). Only three documents have a higher number of self-citations, reaching respectively 18.8%, 34.1%, 22.5%, they are however, in line with ranges reported by Van Raan [vanRaa2001] and Tagliacozzo [Tag1977]. Two of them are written by the same author.

Table 7.- Distribution of citing documents by document type

		Grey literature							Conventional literature			
Docu-ments	Cited documents type	Total number of GL citations	Report	Conf. Paper	Journal article	Theses	Draft	Book and book chapter	Total number of no GL citations	Journal article	Book and book chapter	Conf. Paper
P1	ARTICLE	81	59.3	13.6	13.6	4.9	4.9	3.7	66	93.9	6.1	--
P2	REPORT	54	33.3	18.5	33.3	7.4	--	7.4	32	90.6	3.1	6.3
P3	REPORT	61	52.5	9.8	29.5	1.6	4.9	1.6	22	95.5	4.5	--
P4	ARTICLE	14	64.3	--	14.3	14.3	7.1	--	51	100.0	--	--
P5	ARTICLE	18	66.7	11.1	16.7	--	--	5.6	45	97.8	2.2	--
P6	REPORT	40	62.5	17.5	2.5	10.0	5.0	2.5	8	75.0	25.0	--
P7	REPORT	37	64.9	18.9	8.1	--	8.1	--	13	92.3	7.7	--
P8	ARTICLE	28	64.3	14.3	7.1	10.7	3.6	--	25	100.0	--	--
P9	ARTICLE	35	77.1	11.4	--	5.7	5.7	--	9	77.8	11.1	11.1
P10	ARTICLE	6	--	16.7	33.3	50.0	--	--	33	97.0	--	3.0
P11	REPORT	18	61.1	11.1	27.8	--	--	--	22	90.9	4.5	4.5
P12	REPORT	30	66.7	16.7	3.3	6.7	6.7	--	9	100.0	--	--
P13	ARTICLE	34	52.9	23.5	8.8	11.8	2.9	--	7	85.7	14.3	--
P14	REPORT	31	61.3	32.3	3.2	3.2	--	--	14	100.0	--	--
P15	REPORT	36	58.3	22.2	5.6	8.3	5.6	--	6	100.0	--	--
TOTAL		**523**	**57.7**	**16.3**	**13.8**	**6.3**	**4.0**	**1.9**	**362**	**95.0**	**3.6**	**1.4**

Table 7 shows the distribution of the citing items by document type. GL citing documents represent a complete set of GL document types. "Reports" were the document type most frequently citing our selected documents (57.7%) within GL publication type, while "Journal article" constituted the great majority (95.%) within Conventional literature.

Considering GL document type, citing Reports are homogeneously distributed within our data set, reaching more than 50%, with the exception of P10 and P2 (33.3%). Among the other GL citing documents, GL conference proceedings and GL articles are well represented, albeit with lower percentages than reports (respectively 16.3% and 13.8%). These reach 6.3% and it is interesting to note that even if concentrated in P10 (50%), they cite almost all types of the cited items, disregarding their grey or conventional nature.

6.2. Citation impact

Table 8. – Correlation between publication dates in highly cited and citing documents (FCY, LCY, TCY, Average year and citation peak)

Documents	Document type	Publication year	Total citations	FCY	LCY	TCY	Citation Average/year	Peak
P1	ARTICLE	1999	147	1998	2007	9	16.3	32 (2006)
P2	REPORT	2002	86	2002	2007	6	14.3	24 (2006)
P3	REPORT	1999	83	1999	2007	9	9.2	17 (2004)
P4	ARTICLE	2000	65	2000	2007	8	8.1	14(2003)
P5	ARTICLE	1992	63	1995	2007	13	4.8	8 (2004)
P6	REPORT	1995	48	1997	2006	10	4.8	7 (2000)
P7	REPORT	1994	50	1994	2006	13	3.8	6 (2006)
P8	ARTICLE	1994	53	1994	2006	13	4.0	11 (2006)
P9	ARTICLE	1992	44	1998	2006	10	4.4	7 (2003)
P10	ARTICLE	1999	39	2000	2007	8	4.8	12 (2003)
P11	REPORT	2001	40	2001	2007	7	5.7	11 (2002)
P12	REPORT	1998	39	1998	2007	10	3.9	10 (2005)
P13	ARTICLE	2003	41	2002	2007	6	6.8	11 (2005)
P14	REPORT	1990	45	1991	2005	12	3.7	8 (2004)
P15	REPORT	2003	42	2001	2007	7	6.0	13 (2006)

FCY=First citation year; LCY= Last citation year; TCY=Total citation year

As known many bibliometric studies have claimed that in social sciences a four or five year citation window seems adequate for a paper to be noticed by its scientific community, while in natural sciences the citation window is two years. The latter time period is the one used by ISI to determine the journal Impact Factor, one of the main bibliometric indicators elaborated by E. Garfield (1977). Table 8 relates the publication dates of highly cited documents with those of citing ones. It shows for each cited document the first citation year (FCY) and last citation year (LCY), the total of citation years (TCY) in which highly cited documents receive at least one citation, the average of citations by year. In the last column citation peak is reported. This is the year when highly cited documents receive most citations. 7 documents out of 15 receive citations in the year of publication; most of them (5 documents) are grey, while 2 documents belong to conventional literature. Three documents are even cited in their previous versions: P1 and P13 are cited when respectively available as a GL conference paper and as a technical report, while P15 was already cited as a working paper, before publication in the IMF series *Staff Paper*. For our specific interest in the identification of GL such data highlights once again GL transformation into conventional documents. What is new is that GS is able to track these documents in their early-cited version. The majority of documents (10 documents) in our sample continue to receive citation till 2007. Some documents have a long-term citation life: 7 documents receive citations 10 years after the publication date. This figure does not depend on either the document type or the number of citations received.

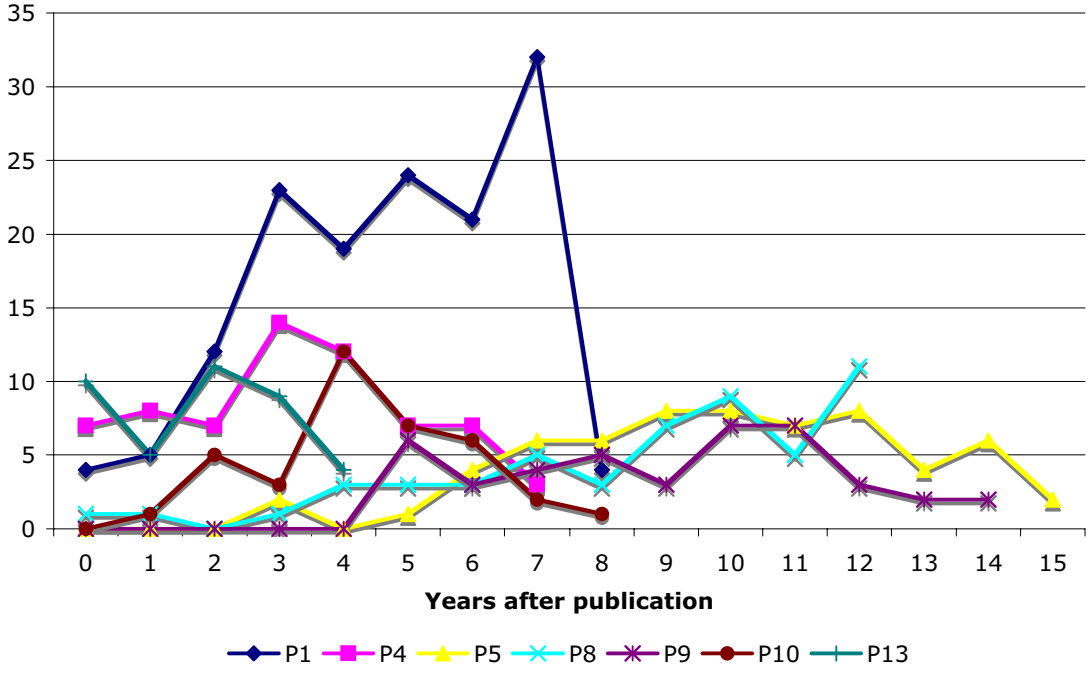


Fig. 7. - Distribution of citations over time in conventional documents

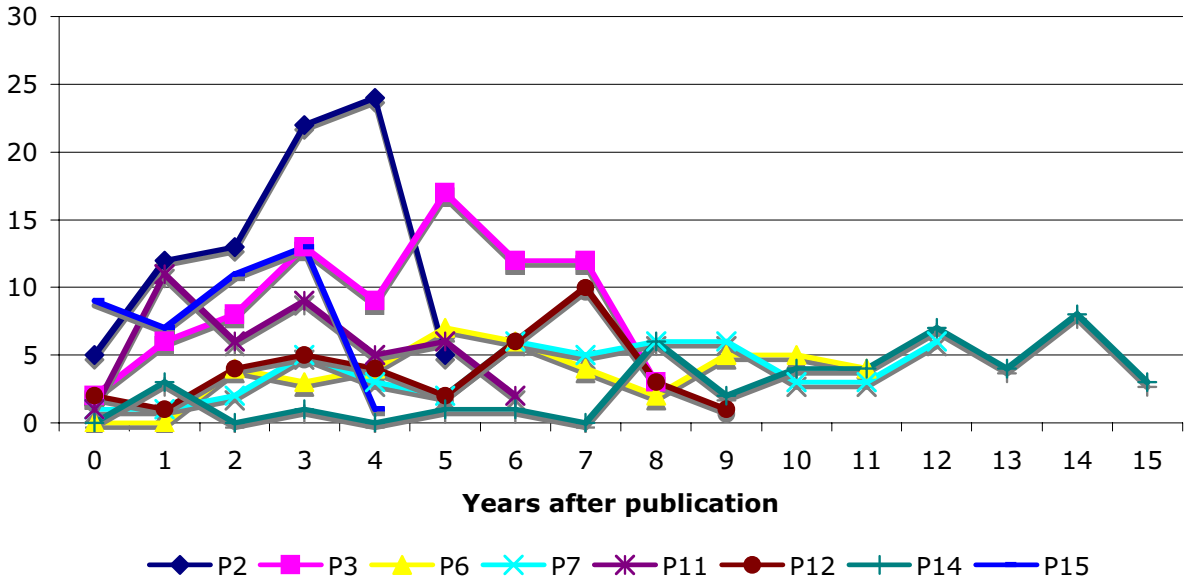


Fig. 8. - Distribution of citations over time in GL documents

The life cycle of the highly cited documents both in conventional literature and GL is represented respectively in figures 7 and 8, given by the number of citation received after publication. Comparing the 2 figures we note that the majority of GL documents receive a higher number of citations than conventional literature and this happens in the first 5 years after publication. Only 3 GL documents have a trend similar to conventional documents. They receive no citations in the first 5 years after publication and afterwards they are noticed by the scientific community and receive citations for a longer period of time, albeit with a lower number of citations. These documents have a long term impact, which is not unusual in demography that is considered a less cumulative field by de Solla Price [Pri1970].

Conclusion

Our first conclusions are related to GS. This citation tracking system can give visibility to GL both as cited and citing document confirming its value in the scholarly communication. Moreover, GS also gives a chance to reconstruct the document life cycle making it possible to retrieve different full text versions of the document, in this way evidencing the GL transformation into conventional literature. Of course there are some limits to using GS: the most important one is that no explicit statement of information sources is given. Nevertheless, it is probable that the development of GS may introduce some changes both in scholarly communication and in the research evaluation activities. On the one hand, scholars might be more inclined to make their scientific production - both GL and conventional documents - available on the web anticipating a wider impact, and on the other, research evaluation should also include GL documents within its assessment activities in order to consider a comprehensive range of scholarly communication.

Turning specifically to the citation analysis, GL documents, and in particular reports, are included in the set of core papers, i.e. in highly cited documents together with well-diffused journal articles. Considering citation counts, citations over time, immediacy impact and self-citations differences between GL and Conventional literature tend to disappear and this may depend on the quality of highly cited documents. Some typical characteristics of GL can be still noticed if we consider that highly cited documents receive more citations by other GL documents and *vice versa*. GL generally receives a higher number of citations in the first 5 years, while conventional literature has a longer citation window (7-10 years).

In the future it would be interesting to further analyse the relationship between cited and citing documents additionally considering other variables such as the corporate source, authorship, country and language. It would also be important to compare our results with other citation databases. In this way we may gain a deeper insight into the diffusion and citedeness of GL in Internet.

Appendix 1. – Highly cited documents

Document	Bibliographic details	Total number of citations	Total number of self citations
P1-A	Ageing of population and health care expenditure: a red herring? / P. Zweifel, S. Felder, M. Meiers. Health Economics, 8 (6) 1999.	147	9
P2-R	World population ageing 1950-2050 / United Nations, Department of economic and social affairs, Population division. New York, United Nations, 2002.	86	0
P3-R	Is the health of older persons in OECD countries improving fast enough t compensate for population ageing? / S. Jacobzone, E. Cambois, J.M. Robine. OECD, Labour market and social policy occasional papers No. 37, 1999.	83	6
P4-A	Population aging: a comparison among industrialized countries / G. F. Anderson and P. S. Hussey. Health Affairs, 19 (3) 2000.	65	0
P5-A	Population aging and the growth of health expenditure / TE Getzen. Journal of. Gerontology: Social Sciences, 47 (3) 1992.	63	4
P6-R	The impact of population aging on savings. Investment and growth in the OECD area / A. Borsch-Supan. Institut für Volkswirtschaft und Statistik, Discussion paper series No. 512, 1995.	48	9
P7-R	The consequences of population aging on private pension fund saving an asset markets / SJ Schieber, J.B. Shoven. National Bureau of Economic Research, Working Paper No. 4665, 1994.	50	0
P8-A	Baby boom population and capital markets / Gurdip S. Bakshi, Zhiwu Chen. The Journal of Business, 67 (2) 1994.	53	1
P9-A	Population aging, social security design, and early retirement / A. Borsch Supan. Journal of Institutional and Theoretical Economics, 148 (4) 1992.	44	15
P10-A	Stable stroke occurrence despite incidence reduction / P. Thorvaldsen. Stroke, 30 (12) 1999.	39	1
P11-R	Apocalypse no: Population aging and the future of health care systems / R.G. Evans, K.M. McGrail, S.G. Morgan, M.L. Barer, C. Hertzman. Social and Economic Dimensions of an Aging Population, Research Papers No. 59, 2001.	40	9
P12-R	Population Ageing and Economic Growth in Seven OECD Countries / Maxime Fougère and Marcel Mérette. Economic Studies and Policy Analysis, Division Department of Finance, Working papers No. 03, 1999.	39	1
P13-A	Labor market effects of population aging / A. Borsch-Supan. Labour 17, Special issue 2003.	41	3
P14-R	Macroeconomic effects of projected population aging industrial countries P.R. Masson. International Monetary Fund, Working Paper No. 5, 1990.	45	1
P15-R	Population Aging and Global Capital Flows in a Parallel Universe / Robin Brooks. International Monetary Fund, Staff Papers, 50 (2) 2003.	42	1
	Total	**885**	**60**

Bibliography

Alberani Wilma, De Castro Paola, (2000). *The use of free literature in health sciences: a preliminary survey results.* Bulletin of Medical Library Association, 78 (4), p. 358-363.

Bakkalbasi Nisa, Bauer Kathleen, Glover Janis, Wang Lei (2006). *Three options for citation tracking: Google Scholar, Scopus and Web of Science.* Biomedical Digital Libraries, 3 (7). Available at http://www.bio-diglib.com/content/3/1/7. (viewed October 2007).

Borgman Christine L., Furner Jonathan, (2002). *Scholarly communication and bibliometrics*. Annual review of information science and technology, 36, p. 3-72. Preprint available at: http://polaris.gseis.ucla.edu/jfurner/arist02.pdf (viewed October 2007).

Cordes, R. (2004) *Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an international Marine scientific advisory body*. The Canadian journal of information and library science, 28 (1).

Cronin B., Snyder H., Atkins H. (1997). *Comparative citation rankings of authors in monographic and journal literature: a study of sociology.* Journal of documentation, 53 (3): 263-273.

Di Cesare Rosa (1994). The evaluation of grey literature impact using bibliometric indicators. The case of Physical sciences. Online information 94, Proceedings of the 18th International online information meeting, London 6-8 December 1994. Oxford: Learned information, p.405-414.

Di Cesare Rosa, Ruggieri Roberta (2006). *The evaluation of grey literature using bibliometric indicators. A methodological proposal.* Proceedings of the 31th Annual conference of the International association of acquatic and marine science libraries and information centres (IAMSLIC) (eds. Kristen L. Anderson and Cecile Thiery). Rome 10-14 October 2005. Fort Pierce FL: Iamslic, p.57-63

Archambault Eric, Vignola Gagné Etienne (2004). *The use of bibliometrics in the social sciences and humanities.* Final report prepared for the Social sciences and humanities research council of Canada. Montreal: Science Metrix

Eugene Garfield (1955). *Citation index for science*. Science v. 122, 108-111

Garfield Eugene (1972). *Citation analysis as a tool in Journal evaluation.* Science v. 178, 472-479

Glanzel, W. (1996). *A bibliometric approach to social sciences. National research performance in 6 selected social sciences areas 1990-1992*. Scientometrics, 35 (3): 193-215

Glanzel, W., Schoepflin U. (1995). *A bibliometric study of ageing and reception processes of scientific literature.* Journal of information science, 21: 37-53

Google Scholar (2007). About Google Scholar. Available at http://www.scholar:google.com/scholar/about. (viewed October 2007)

Hicks Diana (2005). *The four literatures of social sciences*. Journal of management and social sciences, 1(1), p. 1-20.

Hitchcock Steve, Bergmark Donna, Brody Tim, Gutteridge Christopher, Carr Les, Hall Wendy, Lagoze Carl, Harnad Stevan (2002). *Open citation linking. The way forward*. D-lib Magazine, 8 (10), p. 1-25.

Jacso Peter (2005). *Google Scholar and the scientist*. Available at http://www2.hawaii.edu/~jacso/extra/gs/.

Katz J. Sylvan (1999). *Bibliometric indicators and the social sciences. Report prepared for ESRC Polaris House.* University of Sussex.

Noruzi Alireza (2005). *Google Scholar: The new generation of citation indexes*. Libri, 55, p. 170-180.

Pelzer N.L.M.A, Wiese W.H. (2003). *Bibliometric study of grey literature in core veterinary journals.* Journal of medical library association, 91(4).

Price D.J. de Solla (1970). *Citation measures of hard science, soft science, technology and nonscience* , in C.E.Nelson and D.K. Pollok (eds.). Communication among scientists and engineers. Lexington, Ma: Heath Lexington Books, p.3-22.

Tagliacozzo R. (1997). *Self-citation in scientific literature.* Journal of Documentation, 133 (4), p.251-265.

Van Dalen Hendrik P., Henkens Kene (2001). *What makes a scientific article influential? The case of demographers.* Scientometrics, 50 (3) 455-482

Van Dalen Hendrik P., Henkens Kene (2004). *Demographers and their journals: who remains uncited after ten years?* Population and development review. 30 (3), 489-506.

van Raan A.F.J. (2001) *Bibliometric and internet: some observations and expectations.* Scientometrics, 50(1): 59-63

Grey literature on bilingualism in Belgium

Joachim Schöpfel
INIST-CNRS and research groupe "Document Numérique & Usages", France

Abstract
Because of the town Antwerp host of the GL9 conference and because of our own former scientific experience we selected the field of bilingualism for a study on the importance of grey literature in social sciences and humanities. The study is meant to be an additional contribution to comparable scientometric analyses on the distribution of types of publications in different scientific domains.
Bilingualism, the learning and use of two or more languages, is a linguistic, social, educational and psychological reality for many people and most countries. In the heart of Europe, Belgium, a country with two cultural and linguistic populations, with immigration, international business and institutions, is particularly confronted with this reality. Reaction to and part of its multicultural society, Belgium developed since many years a significant interdisciplinary research activity in the field of bilingualism.
The particularity of our study is twofold: First, the research on bilingualism is interdisciplinary, at the crossroad of linguistics, sociology, psychology and educational sciences, each domain presenting its own vectors of publication and communication. Second, while most of the previous studies on the importance of grey literature are citation analyses, our study is based on search results from databases, catalogues, open archives and search engines.

1. On the (relative) importance of grey literature

Grey literature has a role of its own as a means of distributing scientific and technical information, and professionals insist on its importance for two main reasons: research results are often more detailed in reports, doctoral theses and conference proceedings than in journals, they are distributed in these forms up to 12 or even 18 months before being published elsewhere, and some results simply aren't published elsewhere.
But how do researchers use grey literature? One way of evaluating this is to analyze the citations given in their publications. Schöpfel & Farace 2007 provide some empirical evidence based on a synthesis of different scientometric studies (see table 1).

Field	Grey literature citations (in %)
Soil science	14%
Biology	5-13%
Veterinary medicine	6%
Psychiatry (addiction)	1%
Psychology	3%
Engineering Sciences	39-42%
Economics	9-17%
Sociology	7-9%
Education Science	14-19%

Table 1: Part of grey literature in different scientific domains

The relative importance of grey literature is largely dependent on research disciplines and subjects, on methodological approaches, and on sources used. In some fields, especially the life sciences and medical sciences, there has been a traditional preference for conventional distribution media (journals), while in others, such as agriculture, aeronautics and the engineering sciences in general, grey literature resources tend to predominate[i].
A small study of bibliographic references published in four different journals in library and information sciences revealed that grey literature accounts on average around 20% of all sources used. Even so, citations to grey material vary widely between different papers from 0% to 50% and more, depending at least in part on subject areas and methodologies. Most of the non-commercial documents can clearly be identified as unpublished material, doctoral or master theses, proceedings, reports, legal texts, working papers, lecture notes, websites, and even posted messages. But especially for some conference proceedings and other serial resources, the identification of the exact nature of the cited material poses difficulties.

In order to add more empirical evidence on the position of grey literature in scientific publishing, especially in Social Sciences and Humanities, we carry out an alternative approach that is complementary to the citation analyses mentioned above.
We select a scientific field in Social Sciences and Humanities with a specific relationship to the hosting town of the conference GL9 (bilingualism in Belgium), we define a recent time period (2000-2007), we investigate with usual search tools the scientific production (databases, catalogues, search engines, repositories), and finally we evaluate the part of grey literature of this production. Our leading questions were:
(1) What is the scientific production on bilingualism in Belgium?
(2) How important is grey literature for this scientific activity?
(3) What are the particular features of grey literature in this field?
The following section provides a global overview on Belgium and on scientific research on bilingualism, together with some general features related to publications. The third section describes the methodological approach. The 4th section presents the empirical data, the 5th section discusses the limits of the study and resumes the major results, and the 6th section concludes with some apparent features of this sample of grey literature and with two additional questions about quality and access.

2. Bilingualism as an object of scientific study

In our global village, learning and using two or more languages is an everyday reality of many people and most countries. Among the most important reasons for this reality are immigration, international commerce and business, linguistic and cultural minorities. In the heart of Europe, Belgium presents two more particularities: the nation is composed of two cultural and linguistic populations, the French-speaking Walloons in the South and the Dutch-speaking Flemish in the North, while its capital, Brussels, is also the capital of Europe with important international institutions, such as the Commission of European Communities, the European parliament and the NATO headquarter[ii].
Verbal behaviour is linked to economic, social, historical, political, psychological, educational and linguistic factors. Thus, the scientific regard on bilingualism is manifold and depends on the specific background and methodology of the scientist: the psychologist will study the influence of motivation, attitude and cognitive development, the social scientists will focus on identity, role and social group, the linguist will analyse cross-linguistic semantics and syntactic, the educational researcher will evaluate the output of bilingual programs, a psychiatric specialist will publish case studies on pathological verbal behaviour, and a historian will try to understand the cultural, political and economic roots of bilingualism. In many cases, research on bilingualism is conducted by interdisciplinary teams, developing original psycholinguistic, sociolinguistic, psycho-pedagogical or even neuro-linguistic approaches[iii].

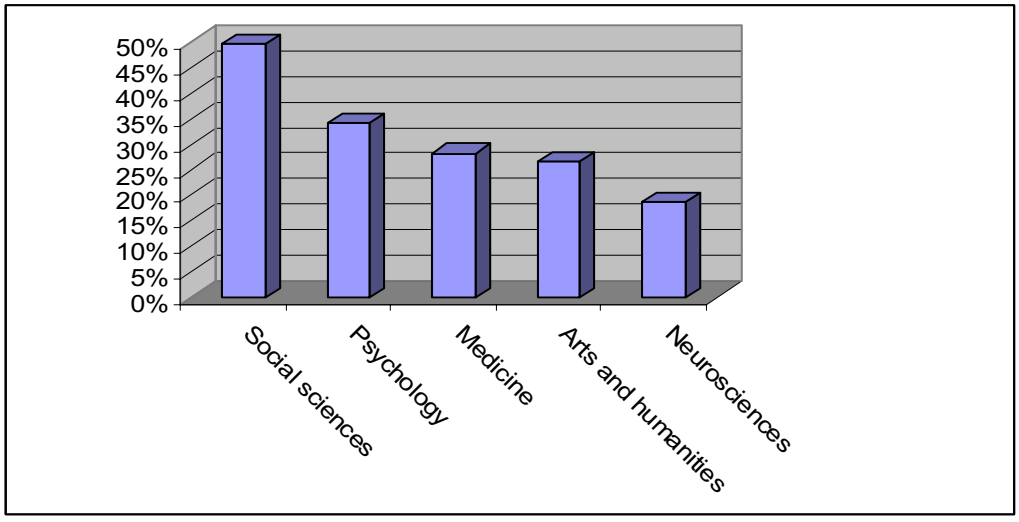


Figure 1: Publications on bilingualism 2000-2007 – scientific domains (Scopus)

50% of recent literature in the Scopus database (from 648 items published between 2000 and 2007) is from social sciences, while psychology, medicine, arts and humanities and neurosciences account for 20-35% (see figure 1).
Figure 1 shows also that articles on bilingualism are partly published in journals indexed in more than one scientific discipline.
In spite of the universal character of bilingualism, only some countries developed a significant amount of scientific research on bilingualism.

Another analysis of the Scopus database reveals that 7 countries produce 85% of the recent literature on bilingualism (see figure 2).
Not surprisingly, more than the half of the indexed literature is published by American or Canadian scientists. All the same, Belgium with 4% ranks at nearly the same level as France (6%) or Germany (5%), which confirms the importance of its scientific production in this field. These results from only one database may not be exhaustive but they seem representative enough.

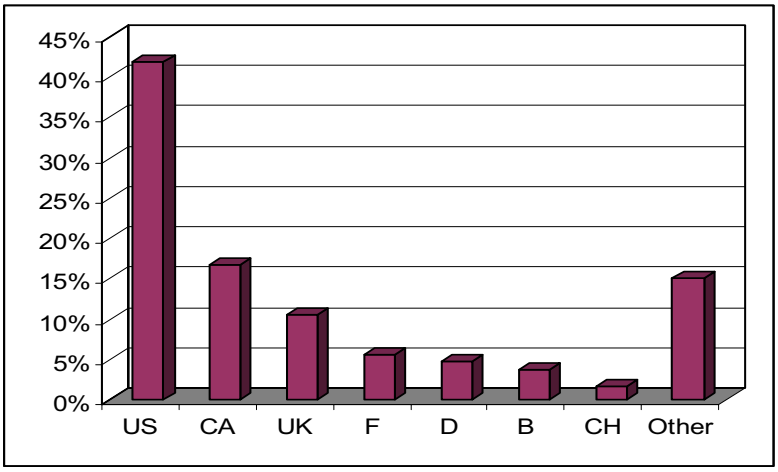


Figure 2: Publications on bilingualism 2000-2007 – countries (Scopus)

3. Methodology

Bibliographic searches in scientific databases and library catalogues and with STI and general search engines were carried out to identify the Belgian scientific production on bilingualism. The following databases were selected:
SCI Science Citation Index (ISI Web of Knowledge)
SSCI Social Science Citation Index (ISI Web of Knowledge)
A&HCI Arts and Humanities Citation Index (ISI Web of Knowledge)
Scopus (Elsevier)
IBSS International Bibliography of the Social Sciences (CSA)
ERIC (CSA)
PAIS International (CSA)
Social Services Abstracts (CSA)
Dissertation Abstracts (ProQuest)
PsycARTICLES (Ovid)
PASCAL (INIST)
FRANCIS (INIST)
ISD (INIST)
Together they cover the core of scientific production (WoK, Scopus, the INIST databases, Diss. Abstr.) and more specific domains related to bilingualism (social science, humanities). The databases were accessed through the national CNRS SS&H portal "BiblioSHS"[iv] or via the local academic campus. Only regret: for technical reasons the *Linguistics Abstracts* database could not be accessed.
The Belgian academic library catalogues that reflect the major national scientific holdings were accessed via the union OPAC *Libis*, via the ILL network *IMPALA* or directly on the universities' web sites (Leuven, Ghent, Brussels etc).
Complementary searches were undertaken with three mostly used generalist and STI search engines:
Scirus (Elsevier)
Google Scholar
Google
The *Directory of Open Access Repositories*[v] (OpenDOAR) was used to identify and search in Belgian open archives. We searched also directly on different institutional web sides (university home pages) for institutional and/or thematic open repositories.
The study was untertaken between May and November 2007.

The search criteria were:

1. Concepts of "bilingualism" in titles, keywords and abstracts (in English, French or Flemish)[vi].
2. The mention of Belgium in titles, authors' address, authors' affiliation or other institutional information.
3. Limitation to recent publications from 2000 to 2007.

The search strategy was rather iterative than linear, alternating between databases and catalogues, repositories and search engines in order to complete affiliations and references, confirm free access and check for other publications from identified authors or co-authors.
Based on the search results two file sheets were created, one with information on authors (first and last name, institution/affiliation, domain and sub-domain), another with information on publications (first author, presence of co-authors, short title, language, year of publication, type of publication, free access on the web, deposit in repository). Data analyses combined information from both tables.

4. Results

The three leading questions for the data analysis are:
(1) What is the scientific production on bilingualism in Belgium? Who publishes in the field of bilingualism, where are the main academic and research structures in this field, which are the dominating topics and approaches?
(2) How important is grey literature for this scientific activity? What is the part of theses, reports and conference proceedings compared to the overall production of results of the research on bilingualism in Belgium?
(3) More specifically, what are the particular features of grey literature in this field? Special attention is paid to the importance of digital information available on the Web, on academic or personal pages or in institutional repositories and other open archives.
Based on empirical evidence, the following sections try to provide at least partial responses.

4.1. Publications: number, year, language and category
Search results in databases, catalogues and repositories and on the web allow for the identification of 164 different publications on bilingualism in Belgium or written by Belgian scientists on this topic, with an average production of 20 publications per year (see figure 3). Considering that 2007 publications are only incompletely recorded in databases and catalogues so far, an overall growth of around 20% can be observed from 2000 to 2007 that is consistent with the general tendency of an annual increase of 2-3% for overall scientific production.

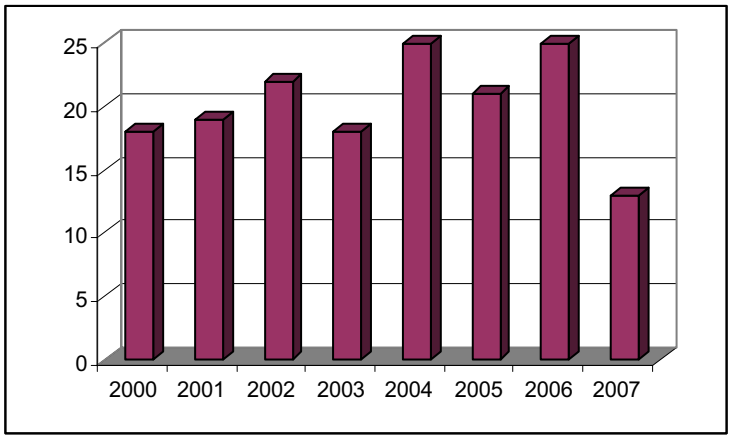


Figure 3: Publication year (sample)

85% of the identified documents are published in English, 13% in French and 5% in Flemish; 2 documents are in Spanish, one in German and another in Italian.
Most of the documents are published as articles in scientific journals (see figure 4).

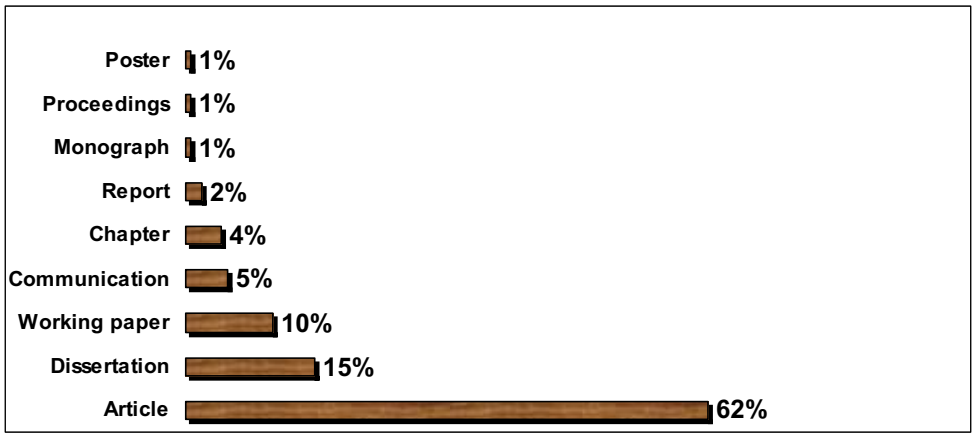


Figure 4: Publication category (sample)

30% are dissertations (e.g. doctoral or master theses), working papers or communications held in conferences. The rest – less than 10% - are book chapters, reports, a monograph, a book with a conference proceedings and a poster.

4.2. Authors, institutions and domains

The 164 publications are written by 81 first authors to whom we added 5 co-authors clearly identified as being affiliated to Belgian academic institutions. Most of the first authors have published one or two documents on bilingualism between 2000 and 2007 but three researchers from the Ghent University occupy a more important and noticeable position: Jan Blommaert with 27 publications, Marc Brysbaert with 9 publications and Robert Hartsuiker with 6 publications.

Affiliation of first author	Number of publications
Ghent University	75
Leuven Katholieke Universiteit (KUL)	13
Antwerp University	11
Brussels Free University (ULB)	10
Lessius Hogeschool Antwerp	5
Kempen Hogeschool	4
Liège University	4
Brussels Research Center for Multilingualism	2
Leuven Catholic University (UCL)	2
Hasselt University	2
Brussels Royal Academy of Science and Arts	1
Namur University	1
Not identified or non Belgian	34

Table 2: Affiliation of first author (sample)

The first authors are affiliated to 12 Belgian scientific institutions, even if some of them are also affiliated to institutions in other countries, such as London University (see table 2).

Obviously the most important Belgian institution for research on bilingualism is the Ghent University with nearly 50% of all publications of the 2000-2007 sample. Four academic campus – Ghent, Leuven (KUL), Antwerp and Brussels (ULB) – cover together 2/3 of the scientific production on bilingualism. 20% of the publications can not be clearly linked to an university or another Higher Education structure or research center, or are written by an author affiliated to a non Belgian institution.

For each publication the major scientific domain was derived from the authors' affiliation, from database indexing and from the topic of publication[vii] (see figure 5).

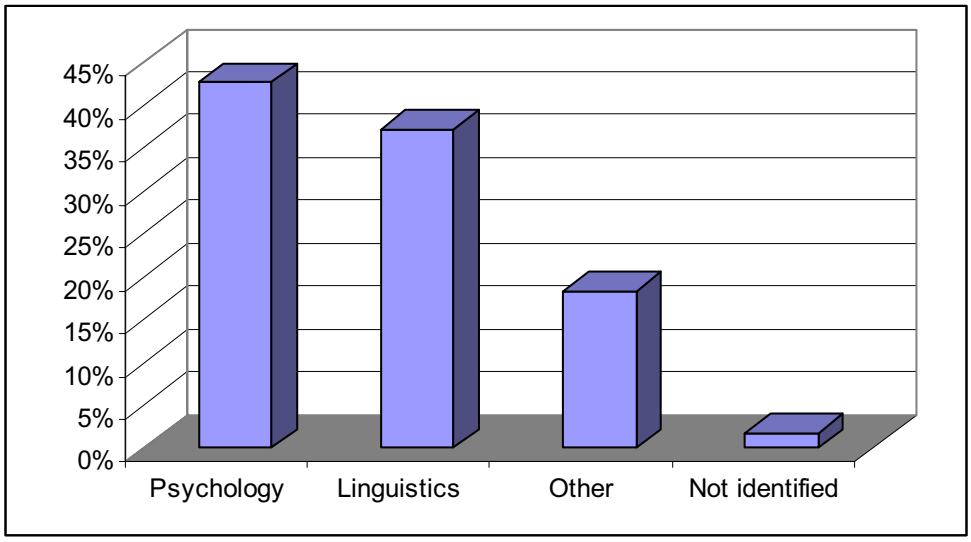


Figure 5: Scientific domains of publications (sample)

Psychology and Linguistics cover together 80% of the sample. The remaining 20% are from other disciplines, e.g. Education, Communication, Neurosciences, Political Sciences, Economics and Business, Literature, History and even Informatics.
Nevertheless, we can distinguish more specific and/or interdisciplinary approaches such as Cognitive Science, Experimental Psychology and Neuropsychology for the Psychology discipline, Ethnography and African Language and Cultures for Linguistics, and Psycholinguistics at the intersection of both disciplines.
Some typical subjects of publications:

- Cross-cultural communication in international schools
- The role of schooling
- Multinational and polyethnic Politics entwined
- Aphasia in a bilingual brain
- Forward and backward number translation
- Ethnography, discourse and hegemony
- Influence of the language-specific orthographic representations
- Bilingual visual word recognition
- Cross-language comedy
- Bilingualism and the development of executive control
- Understanding bilingual memory
- Size of the cross-lingual masked phonological priming
- Cross-linguistic priming
- Universals of language maintenance
- Pathophysiology of language switching
- Multilingualism: an fMRI study
- Investigating narrative inequality

Compared to the international production (content of the Scopus database, see figure 1), there seem to be significantly less Belgian publications on bilingualism in Social and Medical Sciences but as we did not match up our indexing to the Scopus index, this may be at least partly biased by our method.

4.3. Features of grey literature

In our sample, 55 documents (34%) correspond to the criteria of the New York definition of grey literature, e.g. "which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers i.e. where publishing is not the primary activity of the producing body" (GL6, 2004).
Nearly half of these grey documents (45%) are dissertations (master or doctoral theses), another 31% are unpublished working papers, and 15% are communications (see figure 6).

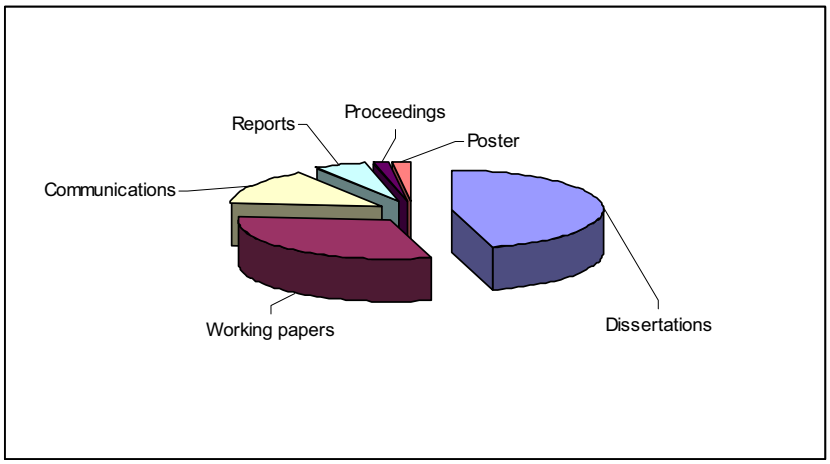


Figure 6: Typology of grey documents (sample)

A closer look on these documents reveals some differences with the "white" publications:
Production: The average number of grey literature is not increasing but stable (around 6-8 publications per year) or slightly decreasing (only 2 identified grey documents in 2005).
Language: The part of non English publications is significantly higher than of articles. While 95% of the published articles and book chapters are in English, the percentage of English grey literature is only 64%, e.g. one third of grey documents are written in French (18%), Flemish (13%) or other languages.
Domains: The part of disciplines other than Linguistics or Psychology is slightly higher (24%) than for articles (18%).
Institutions: In contrast, there seems to be no significant difference regarding to scientific institutions, e.g. no one of them produce more or less grey literature compared to white output.
Authors: The grey documents were published by 38 first authors (47% of all first authors). 24 of them (63%) didn't publish anything else during the observed time period.
Grey to white: Based on the comparison of titles, we can suppose that in at least 8 cases (15%) it exists a direct relationship between grey and white publications, the grey document (working paper, communication, dissertation) preceeding one, two or even four years the publication of an article or a chapter on the same subject.

4.4. Open access to grey documents

For each of the 55 grey documents internet searches were performed, firstly with the mentioned search engines, then with the OpenDOAR search facility, finally directly in some major Belgian academic repositories, especially from the institutions to which the first authors are affiliated. The results are shown in figure 7.

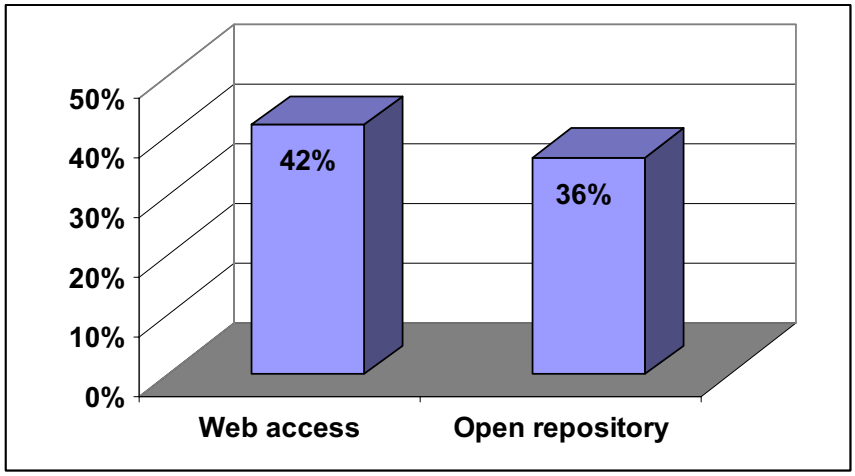


Figure 7: Web access to grey documents (sample)

23 documents (42%) can be accessed on the Web, most of them (20 documents or 36%) are deposited in an open, mainly institutional repository.
Compared to the different categories of grey documents, a tendency or bias can be discerned: In our sample, reports and working papers are more often accessible on the web while dissertations are less.

Maybe that this last tendency is related to our choice to include master theses in the sample that are less often deposited on an academic web sites than doctoral theses. In some rare cases access to documents seems to be lost.

5. Discussion

The presented results are limited by some factors related to the methodology and/or to the subject of the study, e.g.

Language: Belgian academics publish in at least three languages, English, French and Flemish. As a matter of fact, international scientific databases are strongly biased and contain mainly English publications. Even if we tried to counterbalance this bias by searching directly in Belgian catalogues and repositories, searching in Flemish is more difficult than in French or English, even for a German-speaking professional. So some Flemish publications may simply have been overlooked.

Affiliations: Analysis of affiliations was limited to the first author. Nevertheless, 56% of the publications of the sample are co-authored. To gather an in-depth and more comprehensive look on the scientific production of the Belgian academic and research institutions, affiliations of the co-authors should be evaluated.

Scientific fields: Differing from the Scopus database, we only indexed the main scientific domain for each publication which introduces a bias to the major disciplines, e.g. Psychology and Linguistics.

Interdisciplinary: As mentioned above, bilingualism is an object of interdisciplinary research interest. In some "borderline" cases it is not obvious to decide whether a given publication is part of research on bilingualism or should be excluded from the sample, bilingualism being only a marginal topic of limited interest.

The "Belgian aspect" of research: The sample contains (mostly) publications from Belgian authors as well as (some) studies on bilingualism in Belgium published by non Belgian authors. In the second case attention was paid that the publication places this aspect in the centre of interest. This double approach allows a more comprehensive analysis but may introduce some "fuzzyness" in the evaluation of affiliations and open access.

Repositories: Some of the visited institutional repositories provide rather limited search facilities. The alternative approach, browsing or searching "from outside" based on OAI metadata protocol harvesting, may exclude some publications on bilingualism from the sample.

In spite of these limitations, the main results of the study seem to be sufficiently significant and can be resumed as follows:

(1) Scientific production on bilingualism in Belgium: More than 80 academics and researchers published between 2000 and 2007 on bilingualism in Belgium. The annual output is about 20 publications, with a slightly increasing tendency. The main research institutions are the Ghent University and, to a lesser extent, the universities of Leuven (KUL), Antwerp and Brussels (ULB). Most research is performed in Psychology and Linguistics, but the interdisciplinary character of most publications is apparent.

(2) Importance of grey literature: One third of the publications on bilingualism are grey, nearly half of them are master or doctoral theses, followed by working papers and communications.

(3) Particular features of grey literature: Compared to white publications (articles), the part of non English documents is more important. Two third of the authors of grey documents didn't publish their results in articles or books; a smaller part of grey information was published in journals up to four years later. Nearly half of grey publications are freely available on the web, most of the time in institutional (academic) repositories.

6. Conclusion

The goal of our research was to provide additional empirical evidence on the importance of grey literature in Social Science and Humanities, based on a sample of recent publications. The selected topic, bilingualism, is not a strategic or frontline domain such as biotechnology or genomics. Its specificity lies in the interdisciplinary approach, at the frontier of social sciences, humanities and medical sciences and in the strong relationship with the societal environment.

Our sample shows a dynamic and developing scientific research activity and production where the grey literature occupies a significant place, complementary to the articles published in academic journals. Its signification – more than 30% of all publications - can be described by three aspects, timeliness, uniqueness, and community.

Timeliness: One part of the grey documents, in particular dissertations, working papers and communications, present research results one or more years before formal and peer reviewed publications.

Uniqueness: Apparently, one part of the research results presented in grey documents has not been published elsewhere so far, and probably will not be. In these cases, grey literature is the only vector of scientific communication.

Community: Whereas articles in scientific journals are mainly published in English and available for the international scientific readership, grey literature is partly published in the community-specific languages Flemish and French and thus more deeply linked to the Flemish and Walloon scientific communities.
What about quality? Articles published in scientific journals - 62% in our sample – have undergone the usual quality control of scientific information, selection by editors, peer review, revision. Nevertheless, at least 60% of the grey documents have also been evaluated before publication: dissertations are evaluated by academic commissions; and communications have been selected by conference program committees and commented by the public (auditors), followed in some cases by revision before publication. The "quality gap" between grey and white appears thus to be rather relative than absolute, grey literature being at least partly published with added value through scientific peer evaluation.
A final comment should be made on open access to grey literature. The OpenDOAR site records actually (November 2007) 14 institutional, two disciplinary and one aggregating repository located in Belgium, most of them with multidisciplinary and multilingual (English, Flemish and French) content. Even if some of them present limited search functionalities, these repositories appear to be the best way to preserve (archive) and make available grey documents, especially in contrast to hard to find personal web pages or specific academic web sites such as lists of working papers without metadata and/or search facilities. In the future, authors of grey documents as well as scientific and academic institutions should be more conscious of this problem and act in favour of the utility and use especially of institutional repositories as a complementary vector of scientific communication[viii], in particular for working papers, master and doctoral theses and conference proceedings, in order to make and keep them available for a larger scientific community.

7. Bibliography

Correia A.M.R.; Neto M.D. "The role of eprint archives in the access to, and dissemination of, scientific grey literature: LIZA – a case study by the National Library of Portugal". *Journal of Information Science* 2002, 28, 3, 231-241.
Davis P.M.; Connolly M.J.L. "Institutional Repositories". *D-Lib Magazine* 2007, 13, 3/4.
http://www.dlib.org/dlib/march07/davis/03davis.html
Fabbro F. "The bilingual brain: cerebral representation of languages". *Brain and Language* 2001, 79, 2, 211-222.
Farace D.J.; Frantzen J.; Schöpfel J.; Stock C.; Boekhorst A.K. "Access to Grey Content: An Analysis of Grey Literature Based on Citation and Survey Data". In *Seventh International Conference on Grey Literature: Open Access to Grey Resources. GL7 Conference Proceedings, Nancy, France, Dec 5-6, 2005;* Farace D., Frantzen J. Eds.; TextRelease: Amsterdam, 2006.
Hartsuiker R.J.; Pickering M.J.; Veltkamp E. "Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals". *Psychological Science* 2004, 15, 6, 409-414.
Leseaux S.A.; Siegel L.S. "The development of reading in children who speak English as a second language". *Developmental Psychology* 2003, 39, 6, 1005-1019.
Lynch C.A. "Institutional repositories: Essential infrastructure for scholarship in the digital age". *ARL Bimonthly Report* 2003, 226, 1-7.
http://www.arl.org/newsltr/226/ir.html
Portes A.; Schauffler R. "Language and the 2nd generation – bilingualism yesterday and today". *International Migration Review* 1994, 28, 4, 640-661.
Schöpfel J.; Stock C.; Farace D.J.; Frantzen J. "Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit". *The Grey Journal* 2005, 1, 1, 31-40.
Schöpfel J.; Farace, D.J. "Grey Literature". In Bates, M.J. & Maack, M.N. (ed): *Encyclopedia of Library and Information Sciences*. 3rd edition. Taylor & Francis 2007 (forthcoming).
Swenden W.; Jans M.T. "Will it stay or will it go? Federalism and the sustainability of Belgium". *West European Politics* 2006, 29, 5, 877-894.

---

[i] See also Schöpfel et al. (2005) and Farace et al. (2006).
[ii] See for instance Swenden & Jans (2006).
[iii] See for instance some highly cited studies on bilingualism such as Portes & Schauffler (1994), Fabbro (2001), Leseaux & Siegel (2003) and Hartsuiker et al. (2004).
[iv] http://biblioshs.inist.fr
[v] http://www.opendoar.org
[vi] E.g. bilingual, bilingualism, bilingue, bilinguisme, tweetalig, tweetaligheid, second language…
[vii] E.g. title, abstract, author's key words
[viii] See for instance Lynch (2003) , Correia & Neto (2002) and Davis & Connolly (2007).

The use of grey literature in historical journals and historical research : A bibliometric and qualitative approach

Cees de Blaaij
Zeeuwse Bibliotheek, Academic Department

Abstract:
Grey literature is generally accepted as an important part of scholarly communication especially in the "hard" sciences. Since little is known about the use and nature of grey literature in the humanities and specifically in the discipline of history, a systematic analysis was done to characterize the bibliographic references appearing in ten core history journals.
Methods: references from all research articles published in ten core history journals in 2005 were analyzed – five printed and five Open Access history journals - to determine the portion of references from grey literature. The Open Access history journals are free available on the internet. The references were analyzed and categorized according to the type of publication. The uses of grey literature in printed versus Open Access versions were compared.
It will be shown that grey literature as a primary source will be less important than in STM journals. Surprisingly the role of grey literature is of significance in Open Access history journals. The printed history journals show opposed results. Supplementary to grey literature is the development of the format of grey literature as aggregated historical datasets. These datasets are electronically available and will have an important influence on historical research if historians will get more confident in using these sources.

Introduction

The fathers of history - the Greek Herodotus and his peers - didn't care too much about what truly happened in history. Their primary aim was to provide for stories that gave the ancient Greeks a cultural identity through literature or poetry. The literary format was important to legitimize their culture. They didn't think about standards of impartiality and objectivity which are embedded today in the practice of historian scholarship. These standards were the outcome of a process when history became a matured academic discipline. However the research subjects in history were long into the 20th century focused towards "Great Men" and the contextual wars, diplomacy and politics.
Only since the beginning of the 1960's there has been a steady evolution as historians learned from other disciplines and incorporated new innovative techniques and methodologies.
Economic and social science theories challenged existing historical narrative practices. New disciplines like historical demography, social and anthropological history and the use of quantitive methodology[i] opened the way for new perspectives of dealing with the past.
Although some historians felt unhappy leaving traditional narrative paths historical scholarship followed new standardized norms practiced by social science disciplines. The result was increasing differentiation of the discipline history into many sub disciplines. Depending on the type of specialization – methodology, subject and historiographic perspective - the historian's craft is humanities or social science discipline oriented or a mix of both.

Bibliometric characteristics of history

Historical scholarship today offers methods and insights both from the humanities and social sciences. Systematic citation analysis can help us to understand the many qualities and structure of history as an academic discipline. Peters[ii] studied the sources used by historians in the United States who wrote about modern Germany. He asked the question if history should be considered one of the social sciences using the categorization of hardness and softness of a discipline based on three characteristics: the percentages of monographs and serials, the percentage of recent sources, and the percentage of self-citation.
Peters found that the ratio of monographs to periodicals (1.21:1) and the subject dispersion (18.1 percent) showed that history belonged with the social sciences. The percentage of recent references (9.5 percent) was the lowest for any discipline and showed that history should be considered one of the humanities.
These results, when combined, indicate that history is a soft social science that acts as a bridge to the humanities. There is a large variety in the use ratio of monographs to periodicals in historical studies depending on the type of historical scholarship. Traditional narrative oriented historical scholarship relies heavily on the use of monographs just as the rest of the humanities. Citation studies show a predominance of monographs between 60.6 % and 88.5% of the total of citations.[iii] Studies also show if history becomes more interdisciplinary oriented monographs are used less and journals are more favored as an outlet for scholarly communication. In such a situation the use of monographs varies between 36.4% and 64.1%

and the use of journals is represented between 11.6% and 29.8 %.[iv] Nevertheless overall in history – in disciplinary *and* interdisciplinary context - monographs are cited more frequently than journals.
Researchers who have been evaluating scholarly communication agree that there is a lack of enough recently published citation studies[v] for the humanities. For the discipline of history it was suggested that bibliometrics showed itself even to be ahistorical.[vi]
Initially bibliometric research focused itself on the hard sciences. The last two decades more monitoring effort has been made for the social sciences with help from the Social Science Citation Index (SSCI) and for the humanities the Arts & Humanities Citation Index (A&HCI). The humanities arrived late on the citation research scene. Garfield noticed "that humanities scholars are less compulsive about the literature than scientists are."[vii] Perrault noted that researchers in the humanities tend to use older materials than researchers in other disciplines. Currency is not a vital characteristic of humanists' materials and these materials are the "least susceptible to obsolescence"[viii]. On the other hand lack of currency does not appear to devalue significantly the intellectual contribution of a work.[ix]

The use of grey literature in history
Although there has been research on the use of information sources by historians no systematic review till up to now has been undertaken to analyze the use of grey literature as a separate category in historical research. Grey literature is not a well known concept in the discipline of history. Rather the distinction is made between the use of primary and secondary sources.
Primary sources enable the researcher to get as close as possible to what actually happened during a historical event or time period. Secondary sources interpret or analyze a historical event or phenomenon. In this respect history is a house with many rooms: each sub discipline has its own specific ways of dealing with sources. A historical demographer will use different sources than his colleague who is interested in parliamentary politics.

Methods
The use of what we call grey literature in history fits the scope of the definition as formulated during the third International Conference on Grey Literature. Although an exception has to be made in regard to archival records which are used as a source for historical research. Because of their generic nature archival records are not considered to be grey literature as such and are excluded in this study.
To evaluate the use of grey literature by historians this study analyzed and categorized the references of selected historical journals published in 2005.
Several criteria were used to get a representative selection of core journals in the field of history:

1. To assess the use of grey literature relevant scholarly historical journals were selected. The journals contain together a wide variety of historical themes and methods.
2. Selected journals represent not only printed history journals but also Open Access history journals which are electronically available.
3. Selected journals were to be of importance to historians in general on the basis of impact factor and/or widely held by libraries.

Ten journals where selected: five printed journals from the top ten of history journals based on the impact factor compiled by Thomson ISI Journal Citation Reports (JCR) for 2005 and five other open access journals which are available online.[x] Only one of the OA historical journals (Medical History) has an impact factor.

With the advent of new citation databases like Scopus[xi] and Google Scholar new opportunities for better citation research are created in addition to the existing use of Web of Science. A comparative study from Meho and Yang[xii] shows that Scopus and Google Scholar enhance the scope and depth of citation analysis substantially. It is suggested that the new tools "marked the beginning of a new era in citation and bibliometrics analyses".
For finding citations to grey literature in general Google Scholar seems to be the best buy.
GS covers not only print and electronic journals but also a lot of grey literature: conference proceedings, books, theses, dissertations, preprints and technical reports from academic publishers, scholarly societies, government agencies and preprint/reprint repositories from universities.[xiii] GS does have several disadvantages: it does not offer a publishers list, title list, document type identification or any other information about time-span or the refereed status records.
It was also shown that Thomson ISI databases do not count many citations from most conference proceedings.[xiv] However for analyzing the use of grey literature this is an indispensable attribute. Scopus, the citation database from Elsevier, was used to analyze the references from most of the selected journals. The selected printed journals and two OA journals (Medical History and African Studies Quarterly) are indexed by Scopus. The remaining three OA journals were analyzed *in visu.*

Table 1 Core Journals in History

Core journals in History	Number of WorldCat Holding Institutions	JCR Impact factor
Printed Journals		
American Historical Review	2291	1.623
Journal of Modern History	1638	0.568
Comparative Studies in Society and History	875	0.561
Journal of Interdisciplinary History	875	0.312
Social Science History	390	0.286
Open Access Journals		
African Studies Quarterly	313	None
49th Parallel	230	None
History of Intellectual Culture	262	None
Medical History	244	0.333
ERAS	254	None

The OA journals are freely accessible online.[xv] Only six journals out of ten selected as core journals were given an impact factor by Thomson ISI JCR. Only one historical journal has an impact factor of more than 1.0.

All impact factors were derived from the 2005 JCR Social Science Edition. Thomson ISI does not publish a JCR for the humanities. They believe that these are not relevant to literature because the most important publications are books citing other books. Impact factors are relevant where research publications are journal articles that cite other journal articles. Unfortunately only a limited number of history journals has an impact factor but nevertheless some indication of journal relevancy is available to us.
Historical journals have on average a low impact factor compared to the hard sciences. There are several reasons for this. First scholarly communication between historians works differently than in the hard sciences[xvi]. The monograph is the most important vehicle for communication not the journal. Secondly the output in the history field does address an international audience on a different level. Regional studies and local history have a limited range as far as output and as impact so in other words it would be ridiculous to rate such subjects lower than the sciences. Thirdly the absolute value of the impact factor is meaningless. A journal of IF 2 would not be impressive in medicine, while in history it would be top notch. Only the relative value of the impact factor for comparisons between history journals in a mutual context has some relevancy. To supplement the relevancy of a journal in order to select relevant historical OA journals another criterion was added – the number of holding institutions. This represents the availability of OA history journals where an impact factor was missing.

Thomson ISI and Scopus both index the citations of the selected printed journals. Scopus also indexes African Studies Quarterly and Medical History as an OA journal.
ISI Thomson did provide only one impact factor for an OA journal i.e. Medical History[xvii]. Thomson ISI doesn't index citations from historical OA journals.
The selected journals have different approaches to the discipline of history. Three journals represent a social science perspective towards history. Two journals with the highest impact factor reflect traditional historiography. The OA journals have each a distinct approach: geographically, subject and time oriented.

The following steps were taken in the analysis of grey literature in the selected history journals published in 2005:

1. To determine the number of articles that appeared in each issue
2. To determine the total number of references appearing in the research articles
3. To determine the number of references considered to be grey literature.
4. Categorizing the numbers of grey literature references into relevant groups
5. Analysis of the numbers

The types of grey literature were categorized into the following groups: conference proceedings, government publications, news bulletins, organization literature, dissertations and theses. Unpublished dissertations were counted as grey literature. Finally a category miscellaneous was created for those publications which couldn't be categorized elsewhere.

The ten history journals published in 2005 were selected for reference analysis. The journal issues contained a total of 156 articles. A total of 18,366 references were found in the reference lists. It is obvious that history has a very high average of citations per page and citation per article, making it one of the densest fields for citations.

Results

In this study there are two striking differences between printed journals and OA journals. First the average number of references per article is much lower for OA journals than for printed journals.

Secondly it seems as if the use of grey literature is much higher in history OA journals than in their printed equivalent. Use of grey materials varies between 5.2% and a very high 48.4 %. The median for grey materials in OA journals is 13.4% and the median for printed journals with a relatively high impact is 4%.[xviii] An explanation for this could be the different historiographic orientation towards national topics and the treated time period. In general the history OA journals were thematically more oriented towards contemporary modern history and relatively more grey literature was used as a primary source of information.

Table 2 Results for grey literature in History Journals

Journal title	Number of articles	Number of references	Ratio Reference per article	Number of GL references	GL%
Printed History Journals					
American Historical Review	20	4027	201.6	167	4.14%
Journal of Modern History	12	2339	194.9	54	2.31%
Comparative Studies in Society and History	27	2856	105.7	114	3.99%
Journal of Interdisciplinary History	24	4595	191.5	169	3.68%
Social Science History	21	1180	56.2	142	12.03%
Open Access History Journals					
African Studies Quarterly	11	349	31.7	169	48.42%
49th Parallel	9	256	28.4	43	16.80%
History of Intellectual Culture	5	211	42.2	11	5.21%
Medical History	20	2158	107.9	170	7.88%
ERAS	7	395	56.4	53	13.41%

The results show also a strong correlation between impact factor en reference densities for history journals. The citation median for printed journals is 105.7. The median for OA history journals is 42.2. In an earlier combined humanities field studies an average of 62.5 citations per article was found compared with an average of 40.4 citations per article found in a study of biology.[xix] These results are consistent with what was found on citation density in printed and OA history journals.

The use of grey literature appears to be much stronger in the social sciences oriented history journals. The traditional history journals have a higher impact factor but the examined references show much lesser use of grey literature. In one other study where citations from seven journals in English history were examined 12.6 percent could be related to grey literature for the categories unpublished material, manuscripts and theses. Most of the citations - 59.8 percent - were to published, nonserial items; 27.1 percent were to serials.[xx]
In several other studies[xxi] which included categories for types of publication other than books or periodicals the bandwidth ranged from 2.0% to 15.8% for grey material. Manuscripts, dissertations, newspapers, theses, unpublished letters, artifacts, and other miscellaneous sources were included in these categories.

Table 3 Number of grey literature references by type

Journal title	Conf	Gov	Org	News	Diss /Thes	Misc	Total
Printed History Journals							
American Historical Review	9	22	44	14	24	78	191
Journal of Modern History	13	28	4		7	4	56
Comparative Studies in Society and History	8	11	43	5	23	23	113
Journal of Interdisciplinary History	9	20	41	14	25	60	169
Social Science History	14	66	43	4	19	4	150
Open Access History Journals							
African Studies Quarterly		8	37	1	1	12	59
49th Parallel		24	6		1	12	43
History of Intellectual Culture		2	5		4	1	12
Medical History	32	34	66	1	19	18	170
ERAS	2	15	17	3		15	52
Total References	87	230	306	42	123	227	1015
Percent References	6.1	22.7	30.1	4.1	12.1	22.3	100

*Conf =conference; Gov=government; Org=organization; Newsb=newsbulletin; Misc=miscellaneous

A breakdown of grey literature by type of publication can be found in Table 3. Organizational and governmental documents were used in 53.8 percent of references of grey literature. It shows how important they are as a primary source. Government publications accounted for 22.7 percent. There was a wide variety of types of publications due to international themes in the history journals. The geographical dispersion is wide over all the continents. Several of these publications can be accessed online.
The number of documents from non-governmental and not-for-profit organizations is relatively high. Remarkably low however is the 6.1 percent of references for conference proceedings. This is an example of how scholarly communications works differently in the humanities. Conferences play a minor role in history but in other fields – especially the hard sciences – conferences are much more important.
Another interesting result is that dissertations and conference proceedings are much more important in printed history journals than in OA history journals. Perhaps it suggests that OA journals are less serious treated by historians. Several studies noticed that OA Journals aren't well known in the humanities and the there is a lot of uncertainty to embrace electronic journals for quality publishing.[xxii] The printed journals do have a more prestigious image in the humanities and for a number of scholars it is important to be recognized as a scholar.
The high percentage of references in the category miscellaneous reflects the historians working method. They have to recruit evidence from a considerably wider range of sources than scholars in other disciplines. So there is a greater variety in origin, format and language than in other studies. This is not surprising because history as a discipline has no defined body of texts like philosophy, law and literature.

Grey data: the future for history?
Despite the fact that grey literature is lesser used in the humanities than the hard sciences can it be that the influence of grey literature will grow in the future ? In an effort to increase visibility and having access to all kind of documents more and more institutions are producing electronic texts like dissertations, working papers, historical datasets, manuscripts etc. In the Netherlands two projects are relevant for the field of history: the Dare project (Digital Academic Repositories)[xxiii] where Dutch dissertations are available with summaries in English.
Another organization in the Netherlands is DANS: the national organization responsible for storing and providing permanent access to research data from the Humanities and Social Sciences.[xxiv]
The EASY (Electronic Archiving System) System is open to all researchers in the arts, humanities and social sciences. It allows them to store permanently their data and to search data themselves. This includes data from the former socio-scientific Steinmetz Archive and the former Netherlands Historic Data Archive. The objective of EASY is become the electronic repository for data from the social sciences and the arts and humanities. DANS is also a partner in international data organizations. The Netherlands Historical Data Archive comprises a wide range of historical data sets (also known as studies) relating to the Netherlands and its former colonies, and/or created by Dutch historians. The data sets cover the earliest historical periods up to and including contemporary history. At this moment (December 2007) 252 datasets are available. These electronic datasets can give historical research a strong impulse especially in the fields of economic en demographic history. The Dutch National Accounts 1800-1913 and the Dutch Population Census (1795-1971) are a few examples of datasets.
Although the history community is not very confident in using electronic publication channels historians should have more faith in using digital resources and data. This is something which applies to the rest of the humanities as well.

References

[i] Cliometrics is economic specialization within history which uses statistics to measure and quantify elements of an economy. E.g. data have been used to study the economic impact of railways in Europe in the past.
[ii] Peters, Stephen Henry "Characteristics of the sources used by American historians writing on the history of modern Germany". Ph.D. diss., Indiana University, 1990.
[iii] Richard Heinzkill, Characteristics of references in selected scholarly English journals, in: *Library Quarterly*, 50(1), 1980, p. 355; Madeleine Stern, Characteristics of the literature of literary scholarship, in: College and Research Libraries, 44, 1983, pp. 203, 206; Katherine W. McCain, Citation patterns in the history of technology, in: Library and Information Science, vol. 9, 1987, p. 45; Lindholm-Romantschuk, Yeva and Julian Warner, The role of monographs in scholarly communication: An empirical study of philosophy, sociology and economics, in: Journal of Documentation, 52 (4), 1996, p. 395; Cullars, J, Citation characteristics of monographs in the fine arts, in: *Library Quarterly,* 62 (3), 1992, pp. 325-342; Cullars, J., Characteristics of the monographic literature of British and American literary studies, in:

College and Research Libraries, 46 (1), 1985, pp. 511-522; Broadus, Robert N., Use of periodicals by humanities scholars, in: The Serial Librarian 16(1/2), 1989, pp. 123-131.

[iv] McAnally, Arthur M. "Characteristics of materials used in research in United States history". Ph.D. diss., The University of Chicago, 1951; Jones, C., Chapman, M. & Woods, The characteristics of materials used by historians, in: Journal of Librarianship, vol. 4, nr.3, pp. 137-156; Peters, 1990, pp. 43, 62, 74;

[v] Sue Stone (Progress in documentation: Humanities Scholars: Information needs and uses, 1982 p. 292); Jean-Pierre Herubel, Anne L. Buchanan (Citation studies in the humanities and social sciences: A selective and annoted bibliography)
A.J. Nederhof, (Research performances in the Social Sciences and the Humanities, 2006 p. 83)

[vi] Herubel, Jean Pierre V.M., Historical Bibliometrics: its purpose and significance to the history of disciplines in: Libraries & Culture, vol. 34, nr. 4, 1999, p.382

[vii] Garfield, E., Is information retrieval in the arts and the humanities inherently different from that in science? The effect that ISI's citation index for the arts and humanities is expected to have on future scholarship, in: *Library Quarterly*, 50, pp. 43

[viii] Perrault, Anna H., Humanities collection management - an impressionistic/realistic/optimistic appraisal of the state of the art, in: *Collection Management,* 5 (3/4), 1983, p. 4

[ix] Stern, 1983, p. 205

[x] Data Holding Institutions retrieved on Nov 10. 2007 from WorldCat. **Printed Journals:** Jameson, J. Franklin, Henry Eldridge Bourne, and Robert Livingston Schuyler. *The American Historical Review.* [Washington, etc.]: American Historical Association [etc.], 1895. (ISSN) 00028762; *The Journal of Modern History.* [Chicago]: University of Chicago Press, 1929, ISSN 0022-2801; Society for the Comparative Study of Society and History (U.S.). *Comparative Studies in Society and History.* London: Cambridge University Press, 1958, ISSN 0010-4175; Massachusetts Institute of Technology. *The Journal of Interdisciplinary History.* Cambridge, Mass: M.I.T. Press], 1970, ISSN 0022-1953; Social Science History Association. *Social Science History.* Beverly Hills [etc.]: Sage [etc.], 1976; **Open Access Journals**: University of Birmingham. *49th Parallel. An Interdisciplinary Journal of North American Studies.* Birmingham, [U.K.]: University of Birmingham, 1999. <http://www.49thparallel.bham.ac.uk/>; University of Florida. *African Studies Quarterly.* [Gainesville, Fla.]: University of Florida, Center for African Studies, 1997, <http://www.africa.ufl.edu/asq/>; University of Calgary. *History of Intellectual Culture.* [Calgary]: University of Calgary Press, 2001, <http://www.ucalgary.ca/hic/>; Wellcome Trust (London, England). *Medical History.* [London]: Wellcome Trust Centre for the History of Medicine at UCL, 1957, <http://www.ucl.ac.uk/histmed/publications/medical-history/index.html>; Monash University. *Eras.* [Clayton, Vic., Australia]: Monash University, School of Historical Studies, 2001, <http://www.arts.monash.edu.au/eras/editions.htm>

[xi] Product of Elsevier, started in 2004.

[xii] Lokman I. Meho, Kiduk Yang, A new era in citation and bibliometric analyses: Web of Science, Scopus and Google Scholar (working paper, preprint). School of Library and Information Science, Indiana University. Accepted for publication in the Journal of the American Society for Information Science and Technology, 49 pp.

[xiii] Examples: arXiv.org, PubMed, University of Chicago Press, Sage, Springer etc.
See also: K. Bauer, N. Bakkalbasi, An examination of citation counts in a new scholarly communication environment, in: D-lib Magazine, vol 11, no. 9. Retrieved November 25, 2007 from http://www.dlib.org/dlib/september05/bauer/09bauer.html
S. Gardner, S. Eng, Gaga over Google? Scholar in the Social Sciences, in: Library Hi Tech News, 22 (8), pp. 42-45; P. Jacsó, Google Scholar: The pros and the cons, in: Online Information Review, 29 (2), pp. 208-214; A. Noruzi, Google Scholar: The new generation of citation indexes, in: Libri, 55 (4), pp. 170-180..

[xiv] Ibidem, p.2

[xv] Directory of Open Access Journals (http://www.doaj.org)

[xvi] Anton J. Nederhof, Bibliometric monitoring of research performance in the Social Sciences and the Humanities: a review, in: Scientometrics, vol. 66, no. 1, 2006, p.83-84.

[xvii] Medical History (ISSN is a quarterly journal from The Wellcome Trust for the History of Medicine, University College London and is being published in a printed and in a electronic version.

[xviii] The median is used in statistics and bibliometrics to reduce the influence of extreme values.

[xix] Jennifer E Knievel, Charlene Kellsey. Citation analysis for collection development. A comparative study of eight humanities fields, in: The Library Quarterly 75 (2), 2005, p. 147.

[xx] C. Jones, M. Chapman, P.C. Woods, The characteristics of materials used by historians, in: Journal of Librarianship, 4(3), 1972, pp. 137-156.

[xxi] See: note 3

[xxii] Gary Hall, The Impact of the Humanities: or, What's Next for Open Access', Working paper: Conference Berlin 5 Open Access: from practice to impact: consequences of knowledge dissemination, University of Padova, Italy, September 19th-21st, 2007. Retrieved from , <http://www.aepic.it/conf/viewappendix.php?id=270&ap=1&cf=10>

[xxiii] Available online at:<http://www.darenet.nl/nl/page/language.view/search.page>

[xxiv] See also: http://www.dans.knaw.nl/en/over_dans/

Grey literature for development: Some case studies

Bharati Sen
SHPT School of Library Science, SNDT Women's University, Mumbai, India

Grey literature refers to publications issued by government, academia, business, and industry, in both print and electronic formats, but not controlled by commercial publishing interests, and where publishing is not the primary business activity of the organization. Newsletters, reports, working papers, theses, government documents, bulletins, fact sheets, conference proceedings and other publications distributed free, available by subscription, or for sale comprises grey literature (Weintraub 2000) Non-profit organizations and interest groups keep their members informed via a myriad of newsletters and special publications that help shape public opinion. At the turn of the century, documents evolving out of research and development, particularly from the aircraft and aeronautics industries were a very important means of communicating the results of research testing. However, it was the onslaught of World War II, which had the greatest impact on report literature, transforming it into "a major means of communication" (Augur 1989).

The hallmark of that war was the development of technologically-advanced weaponry, from sophisticated tanks to the atomic bomb. These breakthroughs in science made accurate and speedy communications a necessity. The technical report was widely used to disseminate information (White 1984). The decades that followed saw the continuation of staggering amounts of scientific and technological research, which was amassed to improve both military and communication systems.
Generally grey literature has been a major source of information for scientific and technical research and studies. Today however, the importance of grey literature in social sciences and their role in national development is being recognized all over the world. The rise of evidence based policy making in social fields has led to growing 'what works' from existing documented knowledge.

According to Grayson and Gomersall (2003), evidence based policy making (EBP), is a relatively new term. In the policy arena the evidence informed approach embodies several strands of activity, for example the piloting of initiatives, combined with *ex post* or real time evaluation to test their value and effectiveness. However, in recent years there has been increasing emphasis on the review of documented past experience, generally as a guide to 'what works'.
The rationale behind this amassing and assessing of evidence is because:

- There is little that is truly new in this world, either problems or solutions.
- A lot of time and effort is wasted on reinventing wheels.
- Too little is invested in making full use of research findings by placing them into context with other similar studies, and identifying strong messages.
- It is sensible to take advantage of past experience and knowledge, not just for lessons about 'what works' and 'what doesn't work', but 'why and in what contexts'.

A citation study by Ramadevi, Takalkar & Faras (2000) found that the number of citations relating to grey literature and the number of citations of grey references are increasing. Analysis of the articles published in the Journal of Rural Development (India) indicated the importance of and demand for grey literature.

As primary sources grey literature provides un-interpreted, first hand accounts or evidence of an event or experience. These sources contain information or data and are usually written at the time of the event or research. They are usually the original source of information and allow the researcher to analyze a topic without another person's interpretation
The present study is an outcome of the researcher's participation in number of projects that very clearly showed the importance of grey literature in development and the need to organize and create access for future researchers, grass root workers and activists.

Case Study I: Generation of grey literature

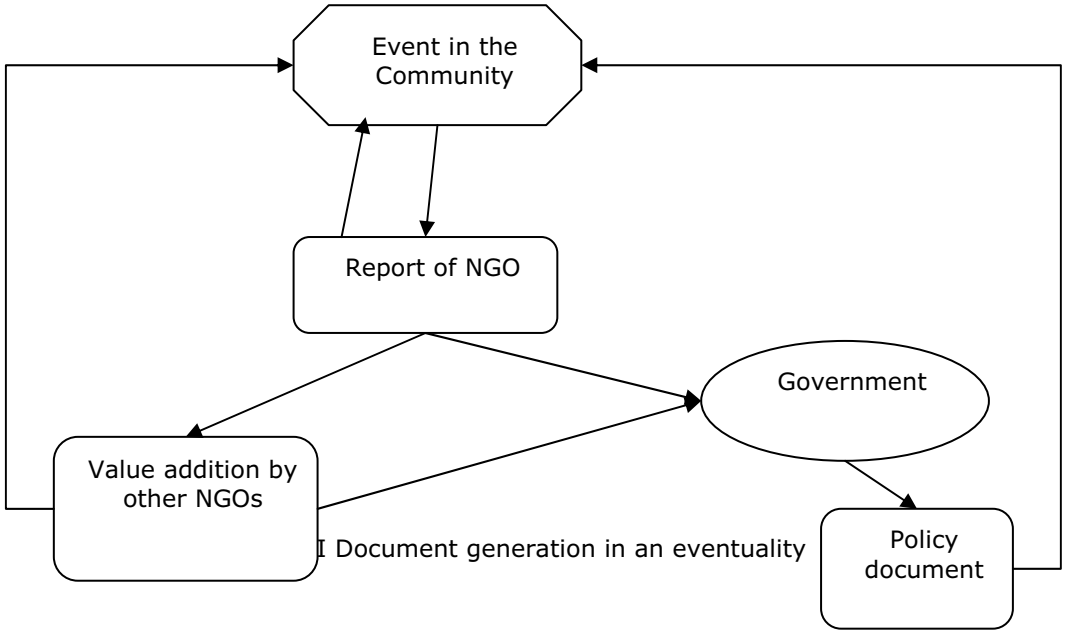
Earthquake in Kutch

In the year 2001 an earthquake measuring 6.9 to 7.9 on the Richter scale struck the district of Kutch in the state of Gujarat, India. The day was 26th January, celebrated as Republic Day in India. Damage to life and property was immense. The rescue, relief and rehabilitation work that followed heavily depended on information of various categories and at various levels of functioning
Being a consultant for a documentation team that worked in Kutch, after the devastating earthquake the researcher found a number of NGOs working .to collect and disseminate information to the stakeholders – information regarding health, housing, education etc. - stakeholders being government agencies and officials at different levels, NGOs and the local community.
Responses to large disaster events are likely to involve a large number of agencies and teams that create severe pressure on requirements for interoperability and cooperation. With numerous agencies wanting to help Kutch, co-ordination became essential. The Civil Society Organizations (CSOs) and corporate houses did not know where to work, how to work. Some corporate houses did not know how to deal with local

people or community. These agencies sometimes had special focus such as health, housing etc. Also Kutch being a large district had long distances to cover. Therefore information and coordination was essential with relief distribution.
For handling the chaotic situation various work and processes had to be defined and formulated. Time was spent on defining the procedures of how to handle the crisis, the organisation of teams the setting up and operation of control rooms etc. The government officials were not at all aware that after the earthquake in Maharashtra on the morning of 30th September 1993 of magnitude 6.3 on Richter scale that rocked the districts of Latur and Osmanabad the Governmant of Maharashtra had appointed a consultant to develop a process document on handling the disaster. This document had established the steps taken for various activities in the first days of the disaster striking.
Later when relief and rehabilitation work was going on a number of local NGOs who had come together under an umbrella organization named Abhiyan having a federated structure decided to make a village gazette for every village of Kutch This was thought to be necessary as many local national and international agencies were working in the area. Many of them had no idea about the local conditions, the caste problems and their implications on relief and rehabilitation work. This gazette would also become a data tool for policy advocacy. Abhiyan from its inception had as its main ideology capacity building, developing long-term solutions.

For any work at the grass root level each agency has to collect information from government office running from one table to another and spend weeks doing so. This process has to be gone through again by any other agency wanting the same data. Finding collated and accurate information at one place is well neigh impossible. Therefore it was decided to set up a data/information bank. With a comprehensive database every organization could immediately get the required information. For example OXFAM had taken up work in two villages of laying pipelines that had been damaged by the earthquake. They required information on water availability and usage in that area, especially the social dynamics whether the Rajputs and the Harijans received equal amount of water etc. The local Abhiyan co-coordinator compiled data about the total number of wells, the number of wells belonging to the different castes such as Darbar, Koli, if there was a pond, whether the water was clean for washing clothes, were Harijan women allowed to use the pond and so on. It was an in-depth study of that locality on ownership of water and its usage. This report ensured that OXFAM project would take care of equitable distribution for the entire community and not get skewed. The report remained with Abhiyan for future use if required (Sen 2007).

A newsletter *Coming Together* started being published. It was not actually a planned publication. It was just a presentation to the Prime Minister when he visited Kutch on 3rd June. Abhiyan tried to get information of as many NGOs as that they could collect. Abhiyan was able to reach 65 organisations. Brief information about each NGO was compiled. The different sectors (such as shelter, health, education, water, livelihood and legal) in which they were working, concerns and recommendations in those sectors were given in bullet points. Gradually, Abhiyan thought of regularising this exercise and the second issue was planned in September. It had more detailed information than the first edition. Not only photographs, maps, showing the extent of damage and work done in different sectors were included, information about how the policies were framed and how, in the initial phase, NGOs had problems in adopting the villages and reasons for delay in plan implementation etc. was incorporated. (For example in shelter sector the reasons for delay in actual implementation of reconstruction). The list of NGOs was enriched with the name of the contact persons, addresses of their Head Offices. This proved immensely helpful for the NGOs who had come to Kutch for the first time. For them it was like a reference book



I Document generation in an eventuality

Recommendation for government policy change was also made through this newsletter.
For example, in the health sector the extent of injury in different areas of Kutch and who/which organisation was doing what type of activity was assessed. This formed the introduction. Next was the status after six months. One of the changes of policy that was required was that compensation to the paraplegics should be more than the compensation for death. Since paraplegic patients were going to be life-long handicapped people they would require regular financial aid for medical treatment through out their lives. So these types of information were also included in the document (Mukherjee 2003).

While the report on process documentation of Latur earthquake remains with the Maharashtra government, the documents generated for relief and rehabilitation in Kutch are available with Abhiyan.

The value of grey documents, and in particular report literature, is increasingly being recognized in social policy fields. For many bodies working in independent research organizations, charities, professional bodies, interest groups, think tanks – it is the primary publication medium

Case Study II: Need for organization and access of grey literature
Gendwaar

Gendwaar doorway to gender information in South Asia (http://www.gendwaar.gen.in/) was created to increase the visibility and enhance access to gender studies information and research in the South Asian Region. Making available catalogues, indexes, bibliographies and directories developed by libraries in the region which have a special focus on gender studies, the website gave links to a growing collection of electronic full text resources from this geographical area. The site had a special section for unpublished literature such as speeches, reports readings etc. Academic resources and activist materials were both included.

While getting the bibliographical details of the published documents did not create any major problems, to be able to gather references let alone full text of 'Report literature' was difficult and time consuming. The site wanted to put up such reports as they embody valuable knowledge that might be applicable in wider contexts. For example, final reports of one project may serve as input to the baseline for another project, or surveys produced for internal decision making may retain their value for some time and have a wider applicability beyond the project that created them.

The institutions felt that their reports were too local to be of any value. In fact those CSOs that have websites do not generally provide a listing of their reports or highlight their findings. (This includes Abhiyan site as on September 4, 2007).

According to Lambert, Matthews, and Jones (ca 2005) the established definition of 'grey literature' needs narrowing for particular contexts, and such a refinement is: *"information produced in a specific working context which is, or might be of value outside that context."* This additional specification is intended to capture the idea that the material, though not intended for publication in the formal sense, may profitably be 'published' to different working contexts.

Social science (especially applied social science or social policy) terminology is diffuse, imprecise and constantly changing. It is frequently 'non-technical' in nature and application, overlapping ordinary everyday language and difficult to distinguish from it.
For the published literature section of the website standard terminologies had been used as descriptors from "Women's Thesaurus'. For Indian words an appropriate term index had been created already. Therefore there was no significant difficulty in incorporating the few documents that became available. They were incorporated into the unpublished section with suitable metadata.

Case Study III: Acquisition storage and access to grey literature
Recommendations of the National Knowledge Commission, India

This section discusses the problems of sharing grey literature with the community from which the information has been culled. Any research report, survey data must enrich the community. This is more crucial for literature that fall into the category of grey literature as they are not available through regular publishing channels.

The researcher had worked on a project involving the women from a rural area. This report will be of some value for future use in decision making. There were no restrictions from the funding agency regarding its publication. Since the report is of limited value regular publishers would not publish it. The researcher wanted to donate a copy to the local library. However local library did not want it, as it was not a regular book. Therefore in the present circumstance there was no way by which local information would become accessible to the local women who would derive some benefit from it.

In an article Thatje, Laudien, Heilmayer and Nauen (July 2007) discussing about the long-term effects of human activities and anthropogenic changes to the environment opined that the loss of (old) data is a major threat to understanding better and mitigating long-term effects of human activities and anthropogenic changes to the environment. They argued that public archiving of original data of present-day research and old (grey) literature and easy public access are important for the future.

In a recent study of indigenous community of North East India it was found that the government of India in its efforts to integrate the tribes was giving them land for agriculture. These tribes had been practicing slash and burn method of cultivation from time immemorial. While the focus of the paper was on forestry and environment one finding was that the distribution of permanent agricultural land has disturbed the movement of elephants in that region as the permanent holdings are fenced and protected.

When the researcher queried the investigator about how her report would be of value for the local community she said that she did not know to whom she could give a copy of the report. For her work also there was no restriction from her funding agency regarding distribution of report copy.

In this scenario the recommendations of the National Knowledge Commission of India (2007) needs to be analyzed. In its report the focus group for libraries recognizes that there are other centres of knowledge in a geographical area and that there is a need to capture knowledge about the local community. Be it a village or a town, libraries need to expand their role to include community information. Libraries should integrate with all other knowledge-based activities in the local area to develop a community-based information system. It is expected that Libraries will be involved in creation and dissemination of area-specific content and relevant information.

In subsequent discussions by the researcher regarding transformation of libraries into community/knowledge centres and in an effort to develop model/s of knowledge centres it became evident that grey literature will be an important component of the collection for information - of the local community, for the local community and by the local community. This should provide greater visibility and accessibility of grey literature for development.

This thought of empowering the local communities is necessary as otherwise they do not actively participate in the decisions for development. This according to Harris (1996) tends to stimulate social exclusion by failing to include local communities.

References

Augur, C. P. (1989). *Information Sources in Grey Literature*. London: Bowker-Saur

Harris, K. (1996). Social inclusion in thee information society. In D. Wilcox (Ed.),

Inventing the future: Communities in the information society. Brighton: Partnership Books.

Mukherjee, N. (2003). *Voices from Kutch – the story of an NGO network: Kutch Nav Nirman*

Abhiyan. Mumbai, Mukherjee KnowledgeWare Association.

National Knowledge Commission (2007). Roadmap of revitalization. Retrieved September 10, 2007, from

http://knowledgecommission.gov.in/downloads/documents/NKC_Library.pdf A

Rama Devi, T., Takalkar, A. (2000). Use pattern of grey literature in rural development.

An Indian case study. *New Frontiers in Grey Literature, Proceedings of the 4th*

International Conference on Grey Literature, GreyNet, Amsterdam, 260-75.

Sen, B. (2007). Information management in Kutch earthquake. In A. Anttiroiko & M. Malkia

(Eds.), *Encyclopedia of Digital Governmen.* (Vol. 2,I p.1041-1046). London: Idea Group Reference.

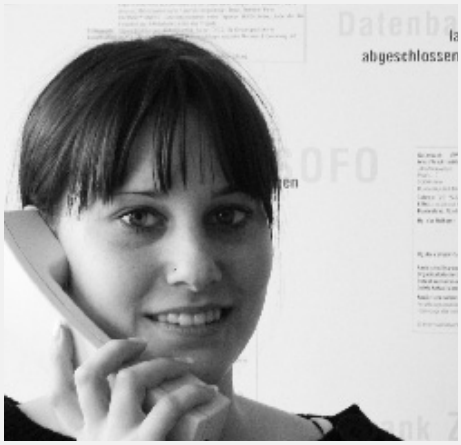
Thatje, S., Laudien, J., Heilmayer, O., & Nauen, C. E. (July 2007). Understanding El Niño — the importance of grey literature in coastal ecosystem research and management. *Marine Policy*, 31 (2), 85-93.

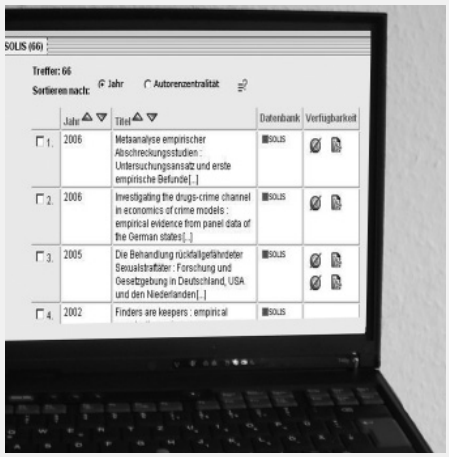
Weintraub, I. (2000). The role of grey literature in the sciences. Retrieved September 10, 2007, from http://library.brooklyn.cuny.edu/access/greyliter.htm

White, H. (1984). *Managing the Special Library.* White Plains, N. Y.: Knowledge Industries Publications, Inc.

gesis
Social Science Information Centre

Services on Grey Literature





... in Eastern Europe

The **gesis Service Agency Eastern Europe** supports East-West cooperation in social sciences and contributes to the networking activities in the European research area. Among other services, the agency offers consultancy on access to grey literature from Eastern Europe.

Access via: www.gesis.org/eastern_europe

... from Germany, Austria, Switzerland

Our database **SOLIS** contains more than 300.000 citations - including about 50.000 citations with abstracts to grey literature in the social sciences in German speaking countries.

Access via:
» www.infoconnex.de
» www.stn-international.de
» www.gbi.de

gesis also provides search services.

More information at: **www.gesis.org**

Digital Documents in Grey Literature: New Challenges

Aleksandr V. Starovoitov
Centre of Information Technologies and Systems of Executive State Authorities, Russia

Lev G. Titarev
International Centre for Informatics and Electronics

Yuri M. Bogdanov and **Leonid P. Pavlov**
The Scientific and Technical Information Centre of Russia

Abstract
The process of transition from paper to digital technology proves to be complicated, expensive and of long duration. Several problems of different nature and scale but equally important – not only technological and technical but also organizational, legal, economic, ethical – have to be solved. Moreover, the problems cannot be solved on corporative level only since they have much to do with the information standard of the scientific community as a whole. The problems of digital document turnover in grey literature are considered in the paper. Among them the need to revise the relevant standards and instructive documents, ensuring the financial support of the digital technologies development, the limited nature of the commercial paradigm in the sphere of scientific and technical information, the problems of copyright and scientific ethics.

Introduction

Several factors result in an increasing importance of grey literature in scientific communications nowadays. A fascinating success of computer networks creating new virtual environment, advanced application software packages for integrating text with complicated mathematical or chemical formulae and illustrative material favour the widespread development of digitally-born documents self-archived by the authors of scientific and technical papers and therefore being related to grey literature. If during several latest hundred years scientific information exchange was founded on printed matter then grey literature becomes the information basis of today's knowledge society. Publication is no longer synonymous to printing but rather may mean presentation on an open Internet site. New digital environment in document preparation, demand and supply shifts accents but presents new challenges to grey literature libraries and information centres.

There is a federal information centre for grey literature in Moscow – the Scientific and Technical Information Centre of Russia (VNTIC) collecting Russian scientific and technical (research and development) reports and dissertations (theses) and disseminating information on them. More than 30 thousand dissertations (approximately 25 thousand candidate and 5 thousand doctoral) and nearly 10 thousand reports arrive at VNTIC annually which means nearly 10 million pages every year. Thus the Centre is quite an information factory continuously processing, storing and making available the grey literature documents to the scientific and scholarly community. Even the quantity of arriving documents in itself turns out to present a problem of quality.

Traditionally VNTIC deals with documents arriving on paper. By the way, one of the reasons for the robustness of paper in bureaucratic environment has much to do with attesting the document by means of head of organization's original signature and official seal. On arrival the bibliographic (the so-called secondary) documents were keyed-in or scanned to be digitally entered into the database and the full-text (the so- called primary) documents were microfilmed to be stored on microfiches. The secondary documents contain metadata (including the abstracts) on primary documents; the corresponding primary and secondary documents have the same inventory number. Since the late nineties practically all the research reports and dissertations have originally been prepared on personal computers. It seemed evident that the scientific community in Russia was ready to present the obligatory documents in digital form. VNTIC had to meet this challenge and started to work out the new technology for digital document collecting, processing, storing and disseminating.

The Federal Scale Problem

The problem of transition from paper to digital document turnover is versatile and many-sided especially in federal centre environment with huge document flows mentioned above, tens of thousands researchers and research organizations engaged in information supply to the Centre in accordance with the Federal Law "On the obligatory copy of documents", about one million potential scientific and technical information

users. The latter figure is based on the following estimation: there are 700 – 800 thousand higher education students and scholars in Russia and around 150 thousand researchers are currently preparing their theses (30 thousand dissertations presented annually by 5 years it takes in average to prepare a dissertation).

Several aspects of the digital problem as applied to the scientific and technical grey literature system on a federal scale should be considered and classified, that is of financial, legal, technological and administrative (organizational) nature. A financial aspect is evident – any new technology needs investments and the conversion of federal collections to electronic form needs federal funding. Legal aspects are diverse and include the necessity to develop new standards, laws, instructive documents and copyright practices. Digital technology suggests the introduction of online operating modes in network environment both for the authors and the users of documents, context search methods and computerized content analysis, computer-aided subject heading assignment, editing and proofreading.

New administrative approaches are supposed to facilitate the complicated problem solution: new forms of interorganizational cooperation help to direct joint efforts at the digital technologies implementation. The authors of this paper represent three organizations, situated in Moscow: Centre of Information Technologies and Systems of Executive State Authorities (abbreviated in Russian as CITIS), International Centre for Informatics and Electronics (InterComputer) and the Scientific and Technical Information Centre of Russia (VNTIC). Under the auspices of the Federal Agency for Science and Innovations of the Russian Federation CITIS, InterComputer and VNTIC have formed a consortium to unite their efforts, resources and experience so that the digital document turnover system in the sphere of scientific and technical grey literature on a federal scale could be realized most timely and effectively.

Challenges and Possible Solutions

Let us consider the main problems that are supposed to be met and solved on the way of going digital following the path of the documents coming in, being inside and going out of the Centre.

Both the primary (full-text) and the secondary (bibliographic) documents should arrive at VNTIC in a digital form through the Internet or on CD-ROMs. The advantages of this approach are evident: it is faster, cheaper, needs neither scanning/keying-in nor spacious operating/storing areas. There are no technical or technological difficulties to implement this mode and relevant software solutions are ready. But several issues of normative and legal nature still have to be solved.

First of all to prevent the arbitrariness in the digital documents presentation a corresponding standard should be introduced. The latest Russian version of the scientific report standard (GOST 7.32 – 2001. System of standards on information, librarianship and publishing. The research report. Structure and rules of presentation. Intergovernmental standard.) was developed within the framework of the Intergovernmental Council for Standardization, Metrology and Certification of the former Soviet states and prepared by the Intergovernmental Technical Committee on standardization MTK191 "Scientific and technical information, librarianship and publishing". Since July 1, 2002 GOST 7.32 - 2001 has been officially acting as a state standard of the Russian Federation. This standard needs urgent revision to reflect the digital document preparation and turnover since it is still oriented on the traditional paper-based publication of reports without the necessary shift of focus concerning the digital representation of documents and corresponding copyright challenges.

The same is true for the normative documents concerning dissertations. The dissertation system is much more centralized and unified in Russia than in the West where the scientific degrees are conferred by universities with a great variety in requirements and quality. In Russia dissertations are presented and collected in a unified way with two federal agencies responsible for the procedures: the Higher Attestation Commission (known in Russian in abbreviated form as VAK) and VNTIC. The legal basis for the VAK activities is the Decision of the Government of the Russian Federation that approves the Regulations of dissertation preparation and presentation. The Regulations are similar to report standards but not so strict and contain the list of obligatory structural components of dissertations, their recommended content and form of presentation. Even the latest VAK Regulations don't address the specificity of preparing dissertations in a digital format.

The second set of normative documents that need revision in view of digital VNTIC arrivals is the VNTIC Instructions on the presentation of the obligatory copy of scientific and technical reports and candidate and doctoral dissertations. The Instructions are approved on the ministry level and registered by the Ministry of Justice of the Russian Federation. The latest version of the Instructions was approved and registered in 1998 so it regulates only the paper documents procedures.

Now that the digital documents have arrived at VNTIC a new group of problems emerges. According to the Federal Law "On the obligatory copy of documents" the functions of VNTIC are twofold: the permanent storage (archiving) of the documents and the dissemination of information on them.

Let us begin with the permanent storage requirement. Of course, when receiving digital documents it would be attractive to get rid of paper and microfilm documents completely. On the one hand, to support the retrospective document collection both on microfiches (which suggests expensive procedures of microfilming and chemical development) and on digital carrier would be prohibitively expensive. On the other hand, permanent storage, that is archiving, requires or so far has required collecting paper or optical copies. There are two possible ways out of the situation. First, to revise the Law in order to delegate the function of permanent storage from VNTIC to some usual archive, not the information centre. Second, to treat the clause of the Law about permanent storage so that it is permitted to have only the digital collection that is retrospectively unlimited digital library of grey literature with all the possible safety measures including storage redundancy and data migration. In any case there is a problem of the retroconversion of microfilmed documents which depends on available financial support only.

The other mainstream activity of VNTIC prescribed by the Law concerns the dissemination of information from its collection. After all, it is keeping the academic and scientific community informed on the achievements in science and technology that the information centres are established for. The efficiency of digitalization is most evident in this direction.

There are several ways and methods of information dissemination from the Centre:
- abstract journals publishing (both on paper and in digital form);
- online search and access to the databases through the Internet;
- full-text document delivery;
- local and distant reading room services.

The only problem with digital technology in publishing is the reluctance of acting editors and proof-readers to change from paper to screen reading. Also, high quality expensive computer monitors should be used for these operations.

All the other dissemination methods face the copyright problem when full-text digital documents are concerned. The straightforward approach of getting permission from and/or concluding agreement with every author is impracticable. In general, the copyright problem has two aspects: commercial (royalty payment) and ethical (plagiarism). It should be perfectly clear that an information centre never sells the right to commercially use the ideas (methods, technologies, materials, etc.) described in the documents of its collection. It just sells copies of the documents compensating for the prime cost of this service.

In case the author may want his/her share in copying income it is most practicable to resort to the help of an intermediary organization like a clearing house. There are at least two organizations in Russia now of the kind of the US Copyright Clearance Center that are engaged in, as they put it, "collective copyright management". The organizations are ROMS (The society of multimedia and digital networks of Russia) and CopyRus (The society of reproduction and copying rights). Both are the members of the International Federation of Reproduction Rights Organizations (IFRRO) and conclude agreements with copying centres and individual authors so that the latter could be paid the relevant fees. Typical terms of the agreement provide for 5% of the centre's copying income going to the intermediary that deducts 20% of this sum while 80% go to the authors.

Also, the digital full texts of reports and dissertations are available to readers on the screens of monitors at the centre's reading room with no means of either paper printout or electronic down-load. Besides the reading room at VNTIC premises, it is advantageous to organize the so called "virtual reading rooms" in many scientific and/or university cities all over Russia allowing distant users to get access to digital reports and dissertations through the Internet. The visitors of the virtual reading rooms are subjected to the same copyright limitations as the VNTIC reading room visitors: they are supposed to be registered as distant VNTIC readers and are allowed to read the texts being within the limits of the distant reading room.

Digital environment stimulates the search of new agreements between authors and users in scientific communication. While trying to optimize the commercial paradigm of author-user relations it seems attractive to change the paradigm radically suggesting that the authors should follow the open archive initiative. That would mean that the authors, say, the authors of dissertations, have nothing against free dissemination of their works including a digital form of the texts.

However, even the open approach does not abolish ethical considerations. Any dissertation or report copy owner must agree to comply with the copyright law and use the copy for research purposes only. In case

of quotations references to the original source must be given. All those fair use principles may seem like something that goes without saying if not the growing scale of plagiarism in the era of digital networks.

The Higher Attestation Commission (VAK) and VNTIC have a long experience of cooperation and involvement in dissertation processing. Now they united their efforts to make the dissertation preparation and presentation process more transparent and public. Earlier this year the Decision of the Government of the Russian Federation has been issued that approved new Regulations for the presentation of dissertations. A really new feature of the Regulations is the system of public notification in advance about the place and time of dissertation presentation, its subject and main results. The technological support and software of the system was developed and now is being implemented at VNTIC.

According to the new rules every dissertation author is obliged to submit an announcement about the dissertation presentation and an extended author's abstract (about 25 pages) of the dissertation in digital form to be put up on a VNTIC Internet server a month before the presentation for candidate and three months before the presentation for doctoral dissertations. The site is open for public so that any interested Internet user is able not only to learn about the dissertation but also to estimate it and take part in discussion about its values or disadvantages. Thus, a sort of voluntary peer review in grey literature is realized.

Concluding Remarks

Nowadays it is universally recognized that digital document turnover in the sphere of scientific and technical grey literature provides both the authors and the public with ample opportunities. However the transition to digital documents on a federal level suggests that many problems of different nature and scale should be solved. Among those organizational, legal, economic, ethical problems are considered in the paper. No matter how many obstacles are on the way it is evident that digital reports and dissertations become an integral part of the knowledge society information basis.

Some Types of Grey Literature: A Polish Context

Marek Nahotko
Jagiellonian University, Krakow, Poland

Abstract
Grey literature consists of documents of very different types. There are some types typical of grey literature, such as scientific and technical reports, and also some types which evoke different opinions. This latter type, such as DŻS documents (documents pertaining to society), standards and patents, can be included in grey literature only because of the specific point of views held in these documents. There are also some specific types of grey literature mentioned particularly by Polish authors and collected by Polish libraries. There are a variety of different historical, economic and social reasons for this. The discussion of these documents' difference from other grey literature begins with a definition of grey literature. For an effective discussion of the different types of grey literature it is important to have an exact definition of the group of literature. A well-known definition was adopted outlining that grey literature consists of those publications "produced at all levels by government, academia, business and industry, both in print and electronic formats, but which is not controlled by commercial publishing interests, and where publishing is not the primary business activity of the organization." Based on this definition, different types of documents can be identified, starting with prohibited (underground) literature, particularly that originating from the 1939-45 period of German occupation and the Communist period. Second, economic-focused materials useful for small business allowing them a quick technology transfer. Third, documents such as unpublished translations into Polish and fourth, articles published in small but locally important journals inaccessible outside of a specific area of the country such as a town or an even smaller locale. Every identified document type includes a description of the document's origin, its role for users, its level of 'greyness,' and its organization and accessibility. Different information systems, both traditional as well computerized, concerned with the types of grey literature outlined above are explored in this paper.
The main goal of this paper is to describe different, and rarely mentioned, types of grey literature examining them from several points of view including the degree to which they fit the grey literature definition, the reasons for their production and use, and the reasons for their local importance. The paper will conclude by highlighting the differences in grey literature types revealed by my research, differences resulting from the varied local needs of non-official communication. Some of the types are stabile and do not tend to disappear. They are usually documents used for technical and economic development. Other documents, resulting from abnormal situations in which there is a lack of legal opportunities to publish social and political, as well as philosophical and historical literature, cease to be developed after the overall situation has changed.

INTRODUCTION

The term 'grey literature' first appeared in scientific publications in the 1970s. Prior to this period, the term was equated solely with reports because it was technical, scientific and economic reports which dominated the materials referred to by label. At the same time, there were also other terms in use to describe the literature, such as 'informal,' 'non-conventional,' 'running away,' 'invisible' or 'half-published.' Attempts to define grey literature are not easy because of the low uniformity of the group of publications. Materials are distinguished mainly by their most important features which differentiate them from other groups. Among the most frequently enumerated features we can find the following common characteristics:

- They are difficult to identify;
- They are difficult to access;
- They are difficult to locate;
- They often come in the form of limited editions;
- They are often inaccessible in bookstores;
- They lack of bibliography registration;
- They are absent in library collections and catalogues;
- They are absent in a publisher's catalogues;
- They are difficult to acquire in libraries;
- They tend be unpublished or published with delay;
- They are rapidly distributed.

In 1997 during the Third International Conference on grey literature, discussions on the most important and distinctive features of the literature were concluded and a generally agreed definition

was reached, which describes grey literature as publications produced at all levels by government, academia, business and industry, both in print and electronic formats, but which are not controlled by commercial publishing interests [Carvalho 2001]. In 1999 at the Washington conference this definition was amended to include the qualification "and where publishing is not the primary business activity of the organization." Currently it is agreed that grey literature is distinguished by the following two criteria: scope, they are usually confined to government, scientific, technological or business publications; and the type of publisher, usually a non-commercial publisher, for whom publishing is not their main form of activity.

We can understand that the form of publication (whether it is traditional or electronic) is not an important or distinctive feature for the purposes of defining grey literature. I wish be clear about this definition of grey literature because it determines the structure of the rest of my paper. Drawing on this well-known definition, my paper will contain a description of some of the different kinds of documents in Poland which can be placed in the category of grey literature. Some features of the document's types will then be compared with features described in the definition, determining that document's level of "greyness" and to what extent it fits the generally accepted criteria.

The definition introduced above allows us to name documents of different types as grey literature. Polish publications often describe some types of document not enumerated in other resources. Based on information made available on the web pages of the Polish Information Processing Centre (OPI) in Warsaw,[i] the system designers identified the following types of grey literature:

1. Scientific, technical, economic, social and other reports developed in national and private institutions.
2. Unpublished conference materials inaccessible via bookstore trade networks.
3. Standards and technical recommendations.
4. Unpublished translations.
5. Papers published in journals issued in a small edition, distributed free of charge, with only local importance.
6. Some official documents.
7. Technical, promotion and advertising documentation.
8. Documents in an electronic form.

The majority of the document types outlined above - i.e. numbers 1, 2, 6, 7, and 8 - are usually named as grey literature and their membership within the literature is typically unquestioned. Other document types - in particular numbers 4 and 5 - are more controversial because their placement on the list is due to the specific needs of the Polish users of the materials.[ii] In Polish professional publications on the subject we can also find some other, even more specific, types of documents which are included in the category of grey literature [Pindlowa, Cisek 1999]. We can, therefore, add the following to our list of grey literature:

9. Grey literature used by small and medium-sized enterprises (MSP), which have a greater importance in Poland in comparison to other countries.
10. Underground (prohibited) literature, published against individuals and regime authorities occupying Polish territory during different periods.[iii]

Drawing on these outlined document types, in the next part of the paper I will focus on four of the specific types of grey literature described above, examining the reasons for their specificity, as well as their place in Polish collections. Each of these document types were chosen because of their different features and usefulness for different groups of users. These characteristics enable these examples to be separated from other types of documents. The chosen document types are: Prohibited (underground) literature, especially originating from the 1939-45 period of German occupation and communist period; Materials useful for small business allowing a faster technology transfer; Unpublished translations into Polish; and finally, Articles published in small journals of local importance, inaccessible outside of a specific town or locale. Every chosen type of grey literature will be described with the use of several predefined criteria. The criteria are as follows:

- The adequacy of the definition;
- Origin;
- Role for users;
- Local importance;
- Level of 'greyness';
- Organization and accessibility.

Prohibited (underground) literature in the context of Polish history

The Adequacy of the Definition

Prohibited literature does not meet the definition requirements exactly as it has not been produced in by government, business or industry. On the contrary, prohibited literature was usually produced to fight against official governmental bodies in Poland. They were not governmental, but rather

anti-governmental publications, and yet in a sense they can be treated as associated with official national institutions. The second part of the definition can be fulfilled without reservations. Prohibited literature was created by political and social organizations, for whom the main aim was to fight for independence, democracy and political changes. Publishing was only a means of achieving these goals. This is why the literature is characterised by a large diversity of political, historical, literary, social and cultural subjects.

Origin

For the purposes of this paper, discussion of prohibited literature will be confined to two periods: the German occupation (1939-1945) and the period of imposed Communist rule (1946-1990). In response to the censorship and political-economic violation common during the German occupation of Polish territory and the later Communist regime, a large number of leaflets, brochures, books and journals were produced and diffused through a system of so-called "second circulation." Publication was limited to a small circle of individuals for reasons of safety and security. An underground Polish state came into being during the German occupation of Polish territory, independent from official German authorities, that undertook secret publishing. Publishing output during the period of occupation is estimated to be about 1500 books and brochures [Kałudzka 2001] and we can divide these publishers into three types: military; political; and civil.

After World War Two was finished, the Polish People's State (PRL) came into power. Described as an independent entity, in reality the PRL was subordinate to the Soviet Union. The resulting, new, antidemocratic orders aroused resistance in Polish inhabitants. To deliver opinions other than those officially approved by the authorities of that time was impossible. After 1946, the Communist Polish state had full control of paper production and distribution, as well as printing and audiovisual equipment [Roliński 2006]. Censorship defined the content and edition of publications and also controlled the final contents of spectacles and films. There were registers of authors prohibited from publication and banned topics of discussion. At the same time, spontaneous publishing activities accompanied every PRL political imposition and were continuously growing. Publishing activities were especially intensive during the events of March 1968. During the second half of the 1970s a rapid growth of independent publications could be observed due mainly to the rise of well organized, illegal, opposition, which led to the later rise of the Solidarity Trade Union.

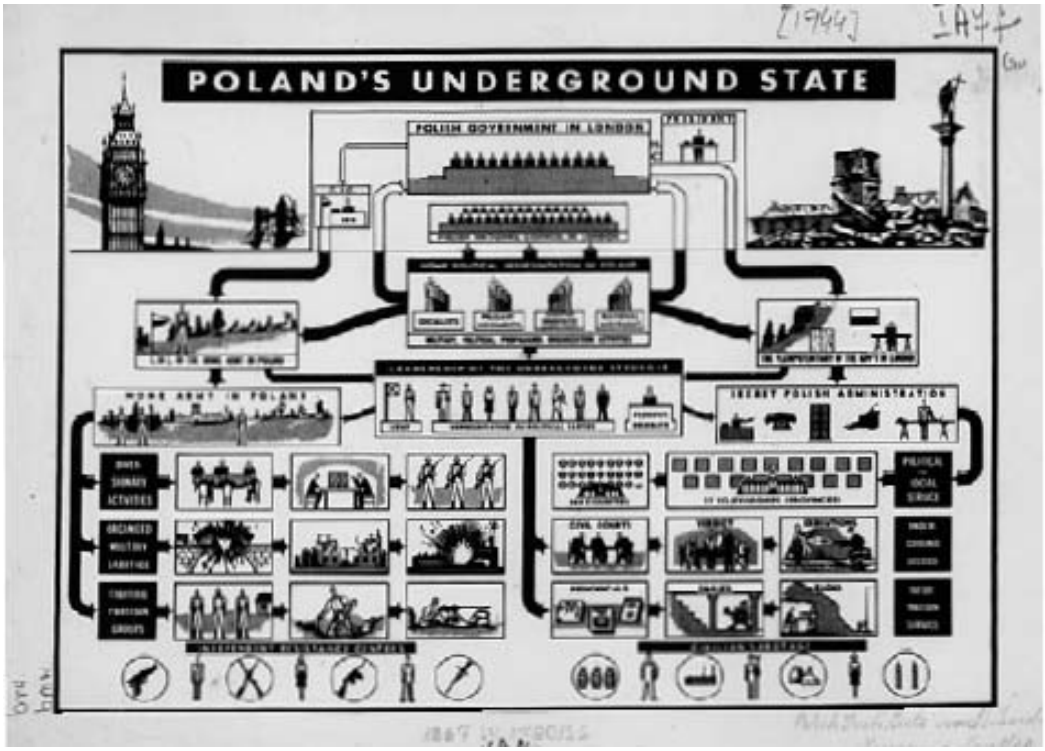


Fig. 1. Leaflet with organizational schema of Polish Underground State [Klamut 2004]

Role for users

Prohibited literature's importance for its users during the period it was published differs from its importance today. During the first period it was useful for actual social and political aims, and it supported progressive changes taking place in periods when it was impossible to make changes to the political system and authorities through democratic methods. In the present these publications hold a more historic value as mementoes of a difficult struggle for independence from alien hegemonies.

Local importance

The materials under discussion here are of bibliophile character. Libraries acquiring such collections are circulating them for scientific and editorial purposes. Because of the rarity of the publications and their deteriorating state (they were produced with the materials to hand, usually of the worst quality) they are stored in specially protected collections.

Level of 'greyness'

There are different forms of prohibited literature. Apart from books, brochures and journals, leaflets, posters, and collections of caricatures were also published. This literature was commonly prepared with the use of a duplicating machine, rather than being printed, and because of this the typography of these documents was completely amateurish. These documents can be rated among grey literature for the reason that neither official nor commercial publishers produced it, consequently they illustrate all the characteristic of grey literature such as difficultly of access and the lack of bibliographic control.

Organization and accessibility

The National Library in Warsaw and the Jagiellonian Library in Krakow hold the biggest collections of prohibited literature. However, neither library possesses a complete set of collections. In the both libraries bibliographies of the publications were prepared. In Warsaw NL a bibliography of underground prints produced during the period 1939-1945 has been developed, and the library also has a central catalogue of the publications. There is also a bibliography of 6500 books unofficially issued during the years 1976 to 1989. The Jagiellonian Library has created a database holding information on "second circulation" publications.[iv] There are quite sizable collections of the literature located in private resources as well.

SMALL BUSINESS IN A POLISH ECONOMIC CONTEXT

The Adequacy of the Definition

The small business application of grey literature is closely related to the field of business and often has additional technical applications, so it easily fulfils the first part of the definition. Also, the paper will examine materials published by non-commercial publishers (often individual authors or teams of authors playing the publisher role), for whom it is not a basic type of activity.

Origin

The source of grey literature for small and medium-size enterprises[v] (MSP) are typically governmental, scientific and industrial institutions as outlined in the definition described earlier in the paper. These institutions develop different types of documents. Some of the most common types of grey literature are research reports, conference materials, doctoral and master's theses, sale and trade catalogues, expert's reports and translations. So many different types of unpublished documents are useful for MSP and all of them are important from their main activity point of view.

Role for users

Grey literature plays a very important role for MSP. MSP entrepreneurs are mainly looking for the following:

- Information about companies, institutions, and organizations. They are looking for information as basic as addresses, level of employment, the names of the board of directors, the organization's economic profile, as well as more complex information such as

a comparison of the company's financial results with one or a number of different business lines;
- Information about markets. Data on changes in domestic and foreign markets, a competitor's promotions and marketing strategy, existing and new products, any potential clients, and any possible development directions;
- Information about finances, i.e. about money and financial instruments;
- Information about products and services, such as patents and trade marks protection, and assistance for monitoring new technical solutions;
- General information about the economy, any political changes which might have an influence on the economy, macroeconomic indicators like the inflation indicator, economic growth indicator, taxes and so on [Cisek 2002].

MSP take many different forms and need a variety of types of advice and mediation. They are largely concerned with information activity and are looking for assistance in the following areas:
- Information on domestic and foreign contracting parties;
- Methods and means of surviving market competition;
- Practical advice.

Most often MSPs make use of textbooks, educational literature, journals and catalogues of different types which offer advice and are useful in solving actual business problems. Business literature which delivers technical information, such as sale and trade catalogues, brochures, and advertising leaflets are also useful. For instance, sale and trade catalogues are utilized by about 90% of companies. In addition, master's and doctoral theses can offer research material for further investigations, especially if the theses include descriptions of a method or sources that are unknown and difficult to access.

Local importance

As mentioned earlier, MSPs fulfil a special role in Poland due a lack of large private enterprises. After more than half a century of a socialist economy, large post-national companies found it hard to cope with a free market economy (see Fig. 2). It is therefore necessary to reconstruct this lost group of enterprises and to ensure the possibility of a rapid growth in their participation level in national economy. Because of the small size of these MSP, firms cannot afford to access expensive traditional publications. Grey literature could be an affordable way for them to access and implement information about new technologies.
One of the main factors enabling MSPs fast development is technical, economic and business information, which contain problem-solving ideas connected with production and enterprise issues in the economy [Strojek 2000]. For several years, we can observe a very fast development of information services and products on the subject, as well as the growth of the potential information users. The main users are MSP, as information obtained at an appropriate place and time is often crucial for their potential development and even survival. The main problem, however, is that there is not sufficient MSP discernment in the local (and foreign) information market.

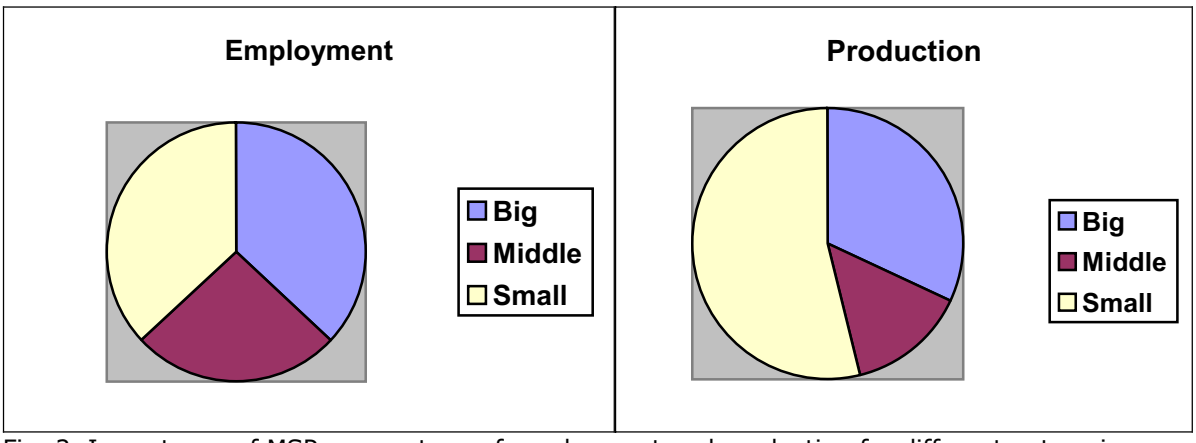


Fig. 2. Importance of MSP: percentage of employment and production for different enterprises.

Level of 'greyness'

MSP uses different types of materials rated among a grey literature, as well as among more traditional publications. The extent to which a document fulfils user needs is, therefore, far more important than a consideration of literature by type alone. Below is a table containing material types used by MSP, together with an indication whether it is grey literature or not.

Materials used during MSP activities	Grey literature Y/N
Research and conferences reports	Y
Doctor and master thesis	Y
Patents and standards	N
Expert evaluations	Y
Translations	Y
Yearly reports of joint-stock companies	Y
Textbooks, handbooks	N
Journals	N
Sale and trade catalogues	Y
Prospectus	Y
Advertising leaflets	Y
Companies address books	Y

Fig. 3. Grey literature for MSP

Organization and accessibility

Demand for grey literature is very large and constantly growing. Entrepreneurs find them a valuable source which helps them to improve their firm's activities and sometimes to survive in a competitive market. This is why it would be useful to develop an information system in Poland about the information delivered to MSP by grey literature. Such a system was planned to be developed by OPI (referred to earlier as POLSIGLE). One of the main aims of the system was to support MSP but unfortunately the project was not fully executed.

UNPUBLISHED TRANSLATIONS IN A POLISH CONTEXT

The Adequacy of the Definition

Here we will look at translations of foreign publications into Polish, but we will narrow the scope of the paper's examination to the areas of technology, economics, organization and the problems encountered by ministers and their dependent organizational units in relation to information activities. It is not about the translations of books or journal papers, standards nor patents. The materials fulfil the requirements of the grey literature definition because they are translations prepared for scientific, technical or business aims. They are also unpublished because the institutions who prepare these translations do not plan to distribute them outside of their organisations.

Origin

Translations are prepared by units of central administration, scientific institutions, foundations and other organizations supporting enterprise development. They are usually prepared by experienced translators who are knowledgeable about foreign languages. In practice scientific workers who have the appropriate language qualifications undertake the task on an individual order basis. Translations are prepared more and more seldom in libraries and information centres because of existence of a large commercial market for such a services, particularly online.

Role for users

Translations in science have decreasing importance because of the common use of English in publications. At the same time, translations have become more important for business, where there is a need to use foreign language documents outlining new technologies, equipment, methods and applications imported from abroad. Translations from Asian languages are especially important. Lack of access to translations may be a significant information barrier for MSP because it then prevents access to the materials themselves.

Local importance

Translations are useful means of breaking information barriers. Currently there are at least thirteen major languages in use by two thirds of the human population. The rest of the world's population use about three thousand different languages including Polish [Chmielewska-Gorczyca, Sosińska-Kalata 1991]. Translations from foreign languages in Poland are important because all

documentation and materials accompanying imported technology and equipment has to be translated into Polish by the producer, importer or user. This is why translations are directly connected with economic development.

Level of 'greyness'

Unpublished translations can be considered typical grey literature publications and they usually are described in such a way in specialist publications.

Organization and accessibility

To the end of 1980s information about Polish unpublished translations was sent to Centre for Scientific Technical and Economic Information (CINTE) in Warsaw, where a special card index was developed. In the past, CINTE has distributed information about translations but unfortunately the Centre was suppressed. Currently there is no information system in existence. The Main Technical Organization (NOT) maintains a list of authorized translators of technical texts. Translations are produced also by specialized firms and they are then the property of the ordering institution. Translations made by human translators prolong the information circulation cycle as it is an especially time-consuming and arduous process. Changes are expected in the field thanks to automatic translation application.

SMALL LOCAL JOURNALS

The Adequacy of the definition

Local journals are low-edition periodicals usually encompassing a parish, housing estate, school, commune, or town district, and addressing a specific, mainly local, social group. They are published by very different bodies including schoolboys, churches and religion associations, social societies, local inhabitants, political parties, industrial plants and other institutions, and persons of different kinds of minorities. In such a differentiated group of publishers we can also find the institutions mentioned by definition of grey literature, like government (local) or business. In practice, the publisher is always non-commercial, a large percentage of the initiatives are not financially self-sufficient (they survive by donations), the others maintain their activities through income gained by advertisements. In a group of local journal publishers 26% of them are independent magazines published by private persons or companies, 36% are local government journals, 22% are parish publications, 10% are produced by societies and foundations, 2% by political periodicals, and 4% are industrial publications (see Fig. 4). The estimated overall amount of local journals in Poland is 2500 titles. Local weeklies amount to around eighty six titles with a total edition of about 530,000 issues. These local weeklies produced outside the eighteen biggest Polish towns have established contacts and cooperation through an advertising agreement entitled Local Weekly. The magazines are independent of local authorities thanks to the payments of their subscribers.

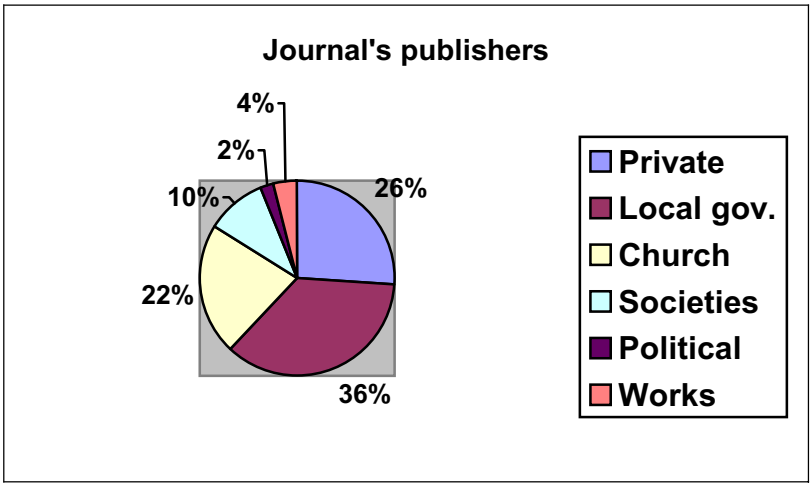


Fig. 4. Local journals publishers

The subject matter of the journals is as different as its publishers. Generally they consist of information interesting to local communities and in particular the community they are made for.

Origin

Local journals appeared in Poland as early as nineteenth century. They were also published in the interwar period and many of them were developed in 1930s. Their most rapid period of development came after 1997 when freedom of speech was restored and also public institutions and authorities of all levels were obliged to present full information about their activities to citizens. The journals were created in participation with the inhabitants of the region, with the most active groups usually being employees of offices, clubs, libraries and so on.. They are developed in places where appropriate intellectual environments exist, which can include small towns and other place but is often near great towns where it is possible to find polygraph services. As a result, the largest number of local journals exists in the most developed regions, areas with a strong industry, knowledge institutions or administration centres (e.g. Warsaw, Krakow, Poznan, and the Silesian Region).

Role for users

There is a social demand for local journals. In small towns there are more users of local journals than all-Polish or regional titles. They satisfy the information needs of local users and offer a view of events from a local perspective, the places where the readers spend most of their lives. They enrich local cultural, and the social, political and professional conditions of an inhabitant's life. They also promote the consolidation of interpersonal ties, helping to develop local people's feelings of worth, as well as assisting with development of so-called small homelands and local patriotism.

Local importance

Local journals play an important role in the local environment. They are often the main source of knowledge about local problems and are an important consultative factor. The market of these publications looks stable with a consistent level of growth. Their basic tasks are:

- Support of versatile, actual and local information;
- Supervising local authorities;
- Promotion of local initiatives;
- Expressing local social opinions;
- Contributing to the integration of local environment;
- Helping to form local opinion;
- The support of local culture and scholarly activities;
- The promotion of small homelands and economic education;
- Providing a forum for advertisements and announcements;
- Entertainment.

Level of 'greyness'

Generally journals are not rated among grey literature. However, local journals have yet to make up a specific group because of the non-commercial nature of their publishers and the limited range of their influence. Both of these features are typical of grey literature. We can decide which local journals successfully meet the grey literature definition by dividing them by their type of publisher (and profile):

Publisher /profile of local journal	Grey literature Y/N
Council journals	Y
School journals (e.g. in high schools)	Y
Religious and church journals	N
Private journals	N
Society (both learned and social organizations journals)	Y?
Political parties journals	N
Industrial plant journals	Y
Economic organizations journals	Y
Minorities journals	N

Fig. 5. Local journals as grey literature

As mentioned earlier, in terms of their contents local journals present a very diverse range of publications including publications in the scientific, technical and social domains.

Organization and accessibility

These kind of publications are quite often acquired and circulated for free by local public libraries. Based on the journals the libraries are documenting contemporary history and development on the lowest level of the administrative division. Thanks to the promotion of local journals carried out by public libraries, the readership of these journals continues to grow. The growth of readership is an important factor in the increasing influence these journals have on local society.

Conclusion

The paper describes four kinds of publications, rated among grey literature, which are especially important for Polish users and rarely created and used in other countries. I do not mean to suggest that these kinds of publications do not exist elsewhere, we can find them in other countries, particularly those with similar history and level of development to Poland. However, the Polish users access to, and use of, these forms of grey literature is of particular importance. The reasons for which users make use of the different forms of grey literature discussed in this paper vary with type. Prohibited literature is important because of its historical significance and importance as archival and bibliophile materials. They testify about the past times of Poland and are numbered amongst grey literature because of how difficult they are to access, and their lack of full information and professional publishers. In contrast, the literature for small businesses is important because of the current needs of of the Polish economy in which MSP play an essential role. Unpublished translations are important in Poland where a non-congress language is in use and, like literature for small businesses, they are important to the economy. Lastly, local journals ensure the continuing development of local societies and democracy on a low administrative level. These journals are rated amongst grey literature for similar reasons to that of prohibited literature, because of their limited access and circulation.

Most of the publication types mentioned here are not completely grey: prohibited literature also consists of reprints from books and journals, which in a normal situation would not be accounted into grey literature; small businesses use different and more traditional types of publications such as standards and patents; and in the case of local journals, journals as a publication type are not usually counted as grey literature. In terms of the publication types discussed by this paper, unpublished translations are the most 'grey' form of literature.

In spite of the important roles these publication types play outlined by this paper, there is no operating information system for any of them in Poland. Earlier existing systems stopped their activity (translations), while other planned systems were not realized (local journals). Some information resources exist for prohibited literature and for MSP literature, yet these resources are dispersed. I, therefore, wish to end my paper by articulating the need for a uniform information system about all types (described here and elsewhere) of grey literature development. In the domestic arena this problem is currently ignored and we cannot count on international information systems to solve the problem for us because crucially those systems are not constructed to satisfy specific local information needs.

Bibliography

Carvalho, Elizabet (2001). "Grey literature and its contribution to knowledge society." In *67th IFLA Council and General Conference* (Boston: August): 16-25.

Chmielewska-Gorczyca, Ewa; Barbara Sosińska-Kalata (1991). Informacja naukowa z elementami naukoznawstwa (Warszawa): 55-57

Cisek, Sabina (2002). Szara literatura jako źródło informacji biznesowej. Zarys problematyki. *EBIB* Nr 11(40) <http://ebib.oss.wroc.pl/2002/40/cisek.php>

Kałudzka, Anna (2001). Cymelia konspiracyjne 1939-1945. *Biul. Informacyjny BN* Nr 4: 35-37.

Klamut, Maria (2004). Kolekcja dokumentów czasu wojny i okupacji (1939-1945) w Zakładzie Dokumentów Życia Społecznego Biblioteki Narodowej. *Rocznik BN* Vol. 36: 213-230

Pindlowa, Wanda; Sabina Cisek (1999). Grey literature as an information source for small and medium-sized business. In: *8th International Scientific and Technical Information in Central and Eastern Europe. Proceedings of the Seminar* (Warszawa, October): 11-12 & 47-53.

Roliński, Adam (2006). O drugim obiegu. http://www.bj.uj.edu.pl/~plok/portal/php/o_obiegu.php>

Strojek, Agnieszka (2000). Znaczenie terminu szara literatura. *Zagadnienia Informacji Naukowej* Nr 1: 64-76

References

[i] See http://www.opi.org.pl for more information. The Polish Information Processing Centre (OPI) was designated as the cooperating centre for the international SIGLE system in a scope of All-Country Polish Information System on Grey Literature (POLSIGLE) design and development.

[ii] Also, the Standards mentioned as the third document type above are often not numbered among grey literature because of their specific features, for example very good organized bibliographic control, which conflict with one of the defined features of grey literature. Also, Standards cannot be treated as specifically polish because the standardisation system in Poland is a part of a global system.

[iii] This paper will focus specifically on the period 1939 to 1990.

[iv] See <http://www.bj.uj.edu.pl/~plok/portal/php/o_obiegu.php> for more information.

[v] In the Polish context a MSP is a company which 1) Employs up to 250 persons (up to 50 is small, between 50 and 250 is medium size) 2) Whose net yearly income from sale of goods, products and services is below 40 million Euros.

Grey Literature in Slovenia – Traditional is solved, what's next?

Primoz Juznic
University of Ljubljana, Slovenia

Abstract:

This paper presents the research done on the special libraries of institutions that are the biggest producers of grey literature in Slovenia. They were defined as the one that do not belong to the formal information and publishing industry (organisations (or individuals working for them) in industry, research institutes, museums and governmental agencies.

The results of the research show how the long-term use and collaboration to the National Union bibliographic system prepared librarians and other information specialists for grey literature handling and dissemination. GL labeled as sensitive material was also found as the important part of some special libraries collections.

Introduction

One of the most widespread used definitions of the grey literature (GL) is the one from the 4th International Conference on Grey Literature (GL4) held in Luxembourg. It says that GL is:
Information produced at all levels of government, academic, 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.
(http://www.textrelease.com/textreleasehome/pressreleases.html)

Grey literature is produced mostly by government agencies, professional organizations, research institutions, universities and other public institutions whose goal is to disseminate current information to a wide audience. The producers of GL are defined as the ones that do not belong to the formal publishing and more generally, information industry. They are mostly organisations (or individuals working for them) in universities and research institutes, and governmental organizations, but also others. GL is often synonymous with unpublished research, although it is not always the case. It is probably easier to describe grey literature through the characteristics of the material, than to define its' content.

So GL is commonly defined as any documents that are not commercially published and is typically composed of research or technical reports, working papers, internal documents, conference (unpublished) proceedings and of course different university dissertations. The greatest challenges involved with these items are the process of identification, since there is limited indexing, and acquisition, since availability is usually marred with uncertainty. Added to this is the absence of editorial control, raising questions about authenticity and reliability.

Traditional grey (or gray) literature is concerned with physical objects, publications produced and distributed by the individuals or organisations that create them. They were not published commercially and usually not indexed by major database vendors. Librarians and other information specialists have invested lot of time and expertise to make these publications available to their clients and other users. If we follow the definition by which the key difference between other sorts of publishing and grey literature is that the latter is not produced as a commercial undertaking, but produced as part of a different communications process, there are more challenges. While the search for these eclectic materials is not new, the development of the Web has increased existing opportunities and open new ones. Trends in communication are changing the notion of grey literature to include also home pages, e-mails, blog postings, wikis etc. The notion of quality has become even more important as grey literature is usually not a subject to the peer review, and must be understood and used accordingly.

World Wide Web sites aid to understanding pf the nature of grey literature by its' various search tools. The focus is upon freely available resources that offer some full-text coverage. While the majority concentrate upon scientific and technical literature, other resources must not be forgotten, as they illustrate the wide range and variation of grey literature.

A comprehensive solution (at least for science and research) would be to integrate access to grey literature within the databases that scholars regularly consult. These databases traditionally identify

peer-reviewed articles. Earlier research has found a certain neglect of GL in these databases. The examination of 11 functioning digital libraries, ten years ago, resulted in a statement that most of the digital library prototypes merely automate existing publishing practices or focus solely on the digitization of the publishing cycle output and do not consider for distribution the large body of " grey literature." Authors suggested methods for expanding the scope and target of digital libraries by focusing on a greater source of technical publications. (Esler, 1998)

This has changed tremendously in last couple of years as now even the most well-known and distinguished database producers include different type of grey literature in their products. Internet has made everything easier. GL is now freely available on many Web sites and is selectively indexed by numerous commercial database vendors. Many organizations and individuals are also providing access to results of their research or other work online. While all these factors present a new optimism, they also raise expectations that everything is available quickly, if not instantly, creating unrealistic perceptions. This gives new challenges also for research that should go beyond the mainstream and often too simplified look on grey literature and Internet/WWW.

Of course here the main disadvantage for non-English speaking countries is the language. Their grey literature is rarely in English, and is usually in their local language, so the interest of these databases is rarely present. So there is still the task of libraries in the first line, to make the access to this type of literature

In her review article D. Luzi (2000) has found a profound change that have appeared in last decades. Studies of GL until the 1990s give evidence to the GL, as something of special interest and awareness of its information value and the need to adopt diffusion techniques and organisational models capable of spreading it to more and more users. They concentrated on the difficulties raised by GL, in terms of acquisition, bibliographical processing and, of course, document retrieval and diffusion, reporting and describing information retrieval sources in detail and identifying libraries and documentation centres where documents could be retrieved.

Next decade the main interest was the booming production of GL and that conventional literature might be falling due to increasing distribution costs. This entails a radical review of roles traditionally performed by publishers and libraries, increasingly called upon to supplement librarianship with the competencies needed to prepare accesses, and develop information packets connecting several types of information and sources. Studies in the 1990s concentrated on these changes, which encompassed fundamental questions of knowledge transfer common to all types of information.

The research in the field of grey literature is becoming an important field of Library and information science. The study using citation analysis followed the work of the authors in the GL-Conference series, as well as authors whom they have cited both in grey and commercial publishing, revived the possibility to follow the work of the authors in the GL-Conference series, as well as authors whom they have cited both in grey and commercial publishing (Schöpfel et al, 2004).

On the other side other bibliometric studies as the one done for the field of nursing and on nursing literature, found out that grey literature was nearly 10% of the citations. The study also revealed an increase in citations to websites, which is anticipated to continue in the future. They proposed further study on the indexing of GL relevant to research use and evidence-based practice in nursing and on how to make this literature easily available to clinicians. (Oermann et al, 2006). Government reports, documents produced by healthcare and other organizations, fact sheets, policy briefs, conference proceedings, and other unpublished documents, in both print and electronic formats, were seen as the most important examples of grey literature.

In other medical field, palliative care, a systematic review was undertaken which has undertaken different steps to grey literature searching. However, their comprehensive search was generally unsuccessful at obtaining unpublished studies. (Cook et al, 2001) As the result they predicted that there will be other avenues for searching grey literature in the near future. These include the new web-based publications and multiple web sites, which should be searched as part of systematic reviewing in the future. It would be interesting to know if this near future has already arrived.

In the survey of the information needs of the public health practitioners, the researchers has searched beyond traditional bibliographic databases to include the diversity of information resources, public health practitioners routinely use or have need to use – GL, conference proceedings, government reports, websites, etc. This methodology has produced a richer and more useful picture of their information needs (Revere, 2007) as GL was an important part of them. Authors have stated that public health librarians and other information specialists can serve a

significant role in helping public health professionals meet their information needs through the development of evidence-based decision support systems, human-mediated expert searching and training in the use information retrieval systems. This is just one of the propositions for the goal of librarians and other information professionals regarding the GL.

Future role of librarians and GL is often seen as a move from functioning as acquisition-oriented reminiscence organisations towards service organisations supporting and facilitating access to digital formed information resources in the networked environment. Librarians have also new opportunities to move towards the production and distribution side of the information cycle.

Libraries and grey literature

Recent research has shed some light on the changes Internet, WWW and Open Access initiative (OI) implied on libraries. A comparative survey of five world major institutions, the document supply of grey literature centres, The British Library (UK), INIST-CNRS (France), KISTI (South Korea) and TIB Hannover (Germany), has found that all of them are more or less deeply involved in the open archive movement, and this involvement has its implications on the policy of acquisition, archiving and supply of grey literature.

All place special emphasis on grey literature and have important grey collections, especially of conference proceedings, technical reports and dissertations. Nevertheless, the relative part of grey document supply differs between them the bigger ones INIST and the British Library with a grey document supply up to 5 per cent, and the CISTI and TIB Hannover with a more important part (.10 per cent) (Boukacem-Zeghmouri, Schoepfel, 2004), so quite a big part of their services.

A case study dealing with one of this world greatest document supplier, the British Library, have revealed the importance of their collection of grey literature and description of their acquiring and supplying grey literature. They stated that during the last 40 years probably the most comprehensive grey literature collection was established and they find out that The British Library will continue to play an active role in preserving and enabling access to this vital resource (Tillett, S. and Newbold, E. 2006),.

Soulouf P et al, (2005) studied how at the University of Rochester (USA) faculty members find, use, and produce grey literature to do their scholarly work. They interviewed 25 faculty members in art and art history, economics, modern languages, linguistics, physics, and political science and all Campus Libraries subject librarians on the topic of grey literature. They found out that subject librarians have a depth of knowledge about the grey literature used in their own disciplines that is extensive, hard won, and valuable. Regarding types of grey literature, theses and dissertations were almost universal types of grey literature, theses and dissertations are almost universal types of grey literature, but also specialized and unusual types of material stand out in some cases.

Research by Ranger (2005) studied access and use of grey literature in special libraries. Thirteen special libraries (government, corporate and specialized academic libraries) located in Texas, Michigan and Washington (U.S.A.). Here interview questions covered different aspect of GL and were answered by librarians. Larger GL collection and use were in academic & government libraries and reports, conference proceedings & newsletters were found to be most used types of GL.

Special libraries in Slovenia

Although the grey literature is a worldwide phenomena, each environment, national included, has its' specific situation and peculiarities. The theme of this paper is to show this as case study in one country and in specific type of institutions, or rather organisational units, **special libraries**.

Special librarianships in Slovenia have a specific history. It is not a long one, actually first list of special libraries was collected in 1965. In 1970 the list contained 147 special libraries (two thirds in industry- as corporate libraries), Reasons for fast growing mainly socio-political. Libraries, later often renamed as INDOK (acronym for information-documentation) centres, were seen as the main vehicles toward important goal. The goal was described as better information of all citizens, named as self-managers, so they can fully fulfil and execute their political role - socialist self-management.

If we leave aside the socialist self-management as a "utopian" system the whole system of special libraries has three main problems:

1. It was build from the top, based on the political programme,
2. There were no information professionals. So most of the new librarians/information specialists were more or less amateurs,
3. Libraries (INDOK centres) had low status in the organisation.

A more then decade was needed to establish more professional approach toward profession, their work and services provided. First professional conference (organized each second year, until today) was organised as late as 1985. A couple of years later the major political changes have taken place, including Slovenia's independence and the change of political system. These have also changed the structures of special libraries. Only the very best in the industry (Krka and Gorenje as the best examples) stayed. Important changes were imposed also on the majority of special libraries elsewhere. In 1993**,** 186 special libraries in Slovenia were on the list, but lot of them were non-active), 2006 list still contains 166 special libraries. Most of special libraries in industry are either closed or non-active. New were mostly established in governmental bodies.

Special libraries are in Slovenia today present in:

- Industry,
- Culture (mostly museums),
- Government (ministries and agencies),
- Science (research institutes),
- Health and medicine,
- Others.

They are mostly smaller libraries with solo librarian. Close cooperation between special and academic libraries have resulted in last two professional conferences organised together. There are huge differences in services, equipment, organisation and staff.

The development and operation of the COBISS system and services (Co-operative Online Bibliographic System and Services) represents the core of the library information system in Slovenia for the last three decades. Almost all libraries and information services within public institutions are part of the COBISS system and their materials part of the database. So Slovenia and its' library system is also very much shaped by its' national online bibliographic system for collecting and making available all the information about libraries collections to all interested users (Seljak, Seljak, 2002). This includes also in some extent grey literature.

It can be said that COBISS serves as a specialised bibliographic instrument to facilitate the identification and retrieval of great part of the grey literature in Slovenia and in Slovenian language. This system has served well in traditional environment, i.e. pre-Web times and is used still now.

Methodology

24 special libraries have been chosen and data about grey literature handling by librarians and other information specialists for grey literature handling and dissemination were collected. A short questionnaire was prepared and library representatives (usually the head of the library) were asked to fill it and also to give their opinion about the subject. Different types of special libraries were chosen to give the overall picture. Special libraries chosen were:

- Two corporate libraries,
- Four governmental libraries,
- Ten special libraries in research institutions,
- Six health and medicine libraries,
- Two special libraries inside museums.

Sample differs from Ranger (2005), as special libraries in Slovenia are somehow differently defined. There are more research institutes and libraries outside the University, most of the ministries and government agencies have their libraries, and health sector has also some very well organised library and information services. We did not include academic libraries in the sample. We included health and medicine libraries because we have found many research in this field (some presented in the introduction), and that gives the impression of the GL importance in the medicine. We have included also museums, as we were interested in possible long time interests in GL with historical-cultural importance.

No one of respondents refused the interview, The main reason was, that the questionnaire was really short, but also lot of the showed the genuine interest in the subject. The results were then analysed. A special carefulness were given to respondents opinions, answers two open questions. Almost all respondents have expressed their opinion on the subject of grey literature. Data collection was dine by the groups of LIS students, as the research was used also as the part of "Special library" course and helping students to learn about grey literature subject.

For consistency's sake, we provided a standard definition of grey literature at the beginning of each interview. We defined it as, ''produced by government, academics, business, and industries, both in print and electronic formats, but which is not controlled by commercial publishing interests and where publishing is -not the primary activity of the organisation".

Questions were following:

1. Do you have any estimation how many documents that could be labelled as grey literature

   originated in your institution? _____

   How much are archived in your library?

   Yearly _____ .

   Together _____

2. Please describe the type of grey literature, which you either archive or make it available.

   _____

   _____

3. How are you cataloguing or documenting these materials:
   a) COBISS,
   b) Other cataloguing system
   c) WWW (library is doing this),
   d) WWW (other organisation unit is responsible)
   e) WWW (authors concern)

4. Lending and supply of the grey literature:
   a) Only to our users
   b) Also to other users but only in the library premises
   c) Lending with no limitations to all users
   d) Everything is on WWW

5. Archiving of the grey literature:
   a) On paper in the library
   b) On paper elsewhere
   c) Only in digital form

6. Please give us your opinion about the importance of the grey literature in the field that is covered by your library.

Results

Interestingly only ten respondents could make an estimation about the number of grey y literature in their organization. Six has no answer to this question,, and the others gave different qualitative or rather descriptive answers. Seventeen libraries has a number about how many are in the library either in actual year or altogether. Others were less precise.

Six, smaller libraries, (research institutes and museums) stated only 10 to 20 documents in their library as grey literature. Eight of them have over 1000 grey literature documents. All of these libraries were part of research institutions, with one industry and health medicine library, that both also have strong research units. Others stated between 10 and 200 grey literature documents. Two

special libraries could not give the precise numbers as they stated that grey literature is the part of the library collection and it is treated as the other library material.

We have already described COBISS (Co-operative Online Bibliographic System and Services), as the system used by most libraries in Slovenia. As COBISS is also connected to SICRIS, system that supports science results evaluation, we have expected that all research institutions used it at least for the documents that can be regarded as the results of scientific research work. This proved right eighteen of libraries (two third), and all libraries in research institutions, use COBISS for cataloguing system and give the data about their grey literature in the Union catalogue of Slovenian libraries. Although some libraries in research institutions have stated that they use COBISS only for the research results materials (they want to be shown in SICRIS - SlovenIan Current Research Information System.

COBISS is the basis of SICRIS. So it is an absolute must, that grey literature with some scientific value finds their way in such a database. This database is the information support of the science system evaluation in Slovenia and is based on citation data, of course mainly from the ISI Web of Science.

Six libraries use other cataloguing systems, either as their primary library automatisation solution or additional one to COBISS. Six libraries are responsible for making grey literature available on WWW pages of the institution, some of the as Intranet solution, other five respondents informed us, that WWW pages and grey literature on it are the responsibility of the other organizational unit. No one gave the responsibility for grey literature to the authors.

There were quite interesting answers regarding lending and accessibility of the grey literature to the potential users. Five libraries limited the access of grey literature only to their users, actually to the employees of their institution. Four more libraries stated that part of their grey literature collection can be used only by their users, the one they labelled as sensitive materials.. As expected both special libraries in industry are in this category. The rest are some research and health and medicine libraries that have limited the use of grey literature too potential outside users. The reason why they have to limit the access, is the importance of these materials for the organization, or the sensibility of information in them.

Majority, fourteen libraries, give access to grey literature (either part of it, or to all) only in the library premises. Mostly these materials are just in one issue, so the concern of not loosing is understandable. Only four libraries said that they lend grey literature without any limitations, another one stated that this is true for the part of these materials, ad other part can be consulted only in library. Interestingly enough, complete availability on the WWW was a case in only one, governmental library and partly reported in another one. It is not a surprise that archiving is almost everywhere in the classic form, on the paper. Actually in all libraries surveyed, except one. Two institutions keep their grey literature in other organisational unit, not in the library, and two have also part of their grey literature collection (new one) in digital form.

The important part of the questionnaire was also the open question, where respondents have expressed their opinion. This shed even some more light on the situation regarding grey literature. Respondents gave us also information about the types of the Grey literature in their organisation. The most common answers were research and similar reports, internal standards and similar documents, congress materials, and quite often, dissertations and other works that originated at the university (diplomas and Master thesis). The latest were actually

Different opinions were also regarding GL materials importance. Some respondents stated them as extremely important, the others tend to find them less important and even predict their importance diminishing.

One of the respondents (health library) had connected GL to open access (OA) initiatives, the subject often discussed in the professional literature. His opinion was that open archives promote the shift from grey to white documents by increasing their availability and access. The boundary between white and grey is shifting, and that we need to know more about grey literature and open access systems in order to update our understanding of what is grey today. This is in line with some other authors in the LIS (Banks, 2004)

Discussion

Results show a rather conservative picture about the special libraries in Slovenia. Their treatment of grey literature is all together, with same exceptions, looks rather classical. This would be a rather superficial conclusion. They are some explanation, which can be derived from the deeper understanding of special libraries work and organization in Slovenia.

- One is the fact that some of them are rather small one (wo)men libraries also working in rather small institutions. In historical part of the introduction we have already pointed this.
- The other reason is COBISS cataloguing system so well rooted in the everyday work and services of the most of the Slovenian libraries and also basis of SICRIS, the system that is the main database used to evaluate the quality of the scientific research work due to their results, publications, where also grey literature is included.
- Third reason is our sample. We have tended to choose libraries that would give us representative picture. On the other side having special libraries in industry and in the research institutes that are doing projects for the industry has to have a certain effect on the answers.

Although results also support the observations that librarians today are far better able to integrate the new grey literature because it is more widespread, and they are accustomed to the added value it offers. It is easier to identify and libraries have come to routinely integrate GL information in the subject-based collections, rather than continue to segregate information in separate units (Gelfand, 1998).

Next step - the challenge of managing new electronic collections that also required substantial new training, charting new policies and procedures in organizations, cannot be regarded as the predominant on, although it was found in some libraries.

Many scholars predict, that grey literature will become indistinguishable from non-grey literature (Marcus, 2006). Valuable grey literature languishes unnoticed and this will not be a case in the nearby future. A common means of increasing access to grey literature is the development of portals to this literature. University produced grey literature is the best example and easiest to put in such portal.

There is unexpected high number of university production (diplomas and dissertations) in Slovenian special libraries. The main reasons are that Slovenian universities are late in institutional repositories where all these material will be available and then special libraries wouldn't have to include those materials in their collections. They are several initiatives, on the different university departments mostly, how to make this type of GL available on the WWW. The start of the National and University library on digital library should be also mentioned (http://www.dlib.si/dlib_eng.asp).

Although expecting too much from the Institutional repositories might be a mistake. Reports from the well developed and available university institutional repositories but largely under populated and underused by its faculty, are something to have in mind before putting too much hope and effort in them. Many researchers use alternatives to institutional repositories, such as personal Web pages and disciplinary repositories, which are perceived to have higher community salience than one's affiliate institution. Many reasons were given for not using repositories: redundancy with other modes of disseminating information, the learning curve, confusion with copyright, fear of plagiarism and having one's work scooped, associating one's work with inconsistent quality, and concerns about whether posting a manuscript constitutes "publishing", says report from Cornell University (Davis, 2007).

The roles of librarians are changed in the process of building the institutional repositories and there are extensions of existing roles in terms of system evaluation, and reference services and may even include content conscription and interpreting publishers' policies. We have found in some Slovenian special libraries a certain type of extremely important grey literature, which will never get the adequate exposure. Different kinds of special libraries have different goals regarding GL, our findings proved to contain a lot of variance among libraries. Surely we must also have in mind that part of GL will never be completely publicly available, but important is that it is still the responsibility and task of librarians and other information professionals, to make them available to the users.

References:

1. Banks, M. (2004). Connections between open access publishing and access to gray literature. Journal of the Medical Library Association, 92 (2), 164-6.
2. Banks, M. (2006). Towards a Continuum of Scholarship: The Eventual Collapse of the Distinction Between Grey and non-Grey Literature? In Conference Papers: International Conference on Grey Literature; 2006, 65-70.
3. Boukacem-Zeghmouri, C., Schoepfel, J. (2006) Document supply and open access: an international survey on grey literature. Interlending & Document Supply, 34(3), 96–104
4. Cook Am. Et al (2001) Efficiency of Searching the Grey Literature in Palliative Care. Journal of Pain and Symptom 22(3) 797-804
5. Davis PM., Conolly. MJL: (2007). Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace. D-Lib Magazine, 13(3/4)
6. Esler, SL.,Nelson, ML. (1998) Evolution of scientific and technical information distribution. Journal of the American Society for Information Science. 49(1), 82 – 91.
7. Gelfand, J. (1998) Teaching and exposing grey literature: what the information profession needs to know – examples from the sciences Collection Building. 17(4) · 159–166
8. Luzi, D. (2000) Trends and evolution in the development of grey literature: a review. The International Journal on Grey Literature, 1 (3),106-116
9. Oermann MH et al. (2007) Information sources for developing the nursing literature. International Journal of Nursing Studies, (in the press), Available online at www.sciencedirect.com
10. Ranger, S.L. (2005).Grey Literature in Special Libraries: Access and Use. In Conference Papers: International Conference on Grey Literature; 2005, 128-135.
11. Revere D. et al (2007) Understanding the information needs of public health practitioners: A literature review to inform design of an interactive digital knowledge management system. Journal of Biomedical Informatics 40(4) 410-21.
12. Schöpfel J., Stock, C.,Farace, D.J., Frantzen, J. (2004) Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. The Grey Journal, 1 (1), 31-40.
13. Seljak, M., Seljak, T. (2002) The development of the COBISS system and services in Slovenia. Program: electronic library and information systems, 36 (2) 89-98
14. Sulouff, P., Bell, S., Briden, J. (2005) Learning about grey literature by interviewing subject librarians: A study at the University of Rochester. College and Research Libraries News, 66 (7), 510-15.
15. Tillett, S. and Newbold, E. (2006), "Grey literature at the British Library – revealing a hidden resource", Interlending & Document Supply, 34 (2), 70-3.

How do High-Energy Physics scholars discover scientific information?

Anne Gentil-Beccot
CERN, European Organization for Nuclear Research, Switzerland

Abstract
Grey literature has always been the main conduit of scholarly communication for High-Energy Physics (HEP) researchers. An efficient way of searching and accessing this information is a central part of their research workflow. In 2007, a survey was conducted to understand which information resources HEP scholars use to find the information they need. The results of this survey are presented. Over 2000 answers, representing about one-tenth of the active HEP community, were collected and show that community-driven resources largely dominate the landscape, with commercial services serving only a small proportion of the users. In addition, HEP scholars appear to use different tools for different information needs, which are clearly prioritized. Finally, the results of the survey shed light on the future information needs of HEP scientists over the next five years.

Introduction
The High-Energy Physics community is a small and cohesive community, counting about 20,000 scientists. Grey literature has always been the main vehicle of scholarly communication for this community, which has built specific information systems to access and use these resources. As well as information services specifically developed by the community, some other information services exist: commercial online databases, providing metadata about all scientific literature; journal platforms, serving metadata and full-text of articles in the field; the pervasive Google and its scholarly instance, Google scholar...
These three categories of information systems constitute the current landscape of information provision and search platforms used by the HEP scientists. In addition, online services are changing more and more, new tools are developed and new ways of interacting with users evolve. In this environment it is important to understand what is the current usage of information services[i] in HEP. Do scientists prefer publishers' own portals, or Google, or the field-specific access provided by HEP information resources? Do they expect to see some changes? Which of these platforms is better placed to answer any need for change? What kind of evolution should be envisaged for an information system to meet the needs of HEP scholars?
After describing the specific relation there is between HEP community and grey literature, this article describes the survey jointly conducted in 2007 by several HEP research institutes about the usage of HEP information systems. An overview of the methods used in the survey is first presented, followed by the main results: the large domination of community-based services and an insight into current and future information needs of HEP scholars.

Grey literature in High-Energy Physics
The High-Energy Physics (HEP) community has always used grey literature as its main vehicle of communication[ii]. Half a century ago, before the time of electronic journals, the delay between the submission of a scientific paper and the time when it reached the reader appeared unacceptable to HEP scientists, who embraced preprints as their main means of communication. One of the origins of this phenomenon is the peculiarity of a two-soul community: experimental physicists working at accelerators of subatomic particles of ever-increasing energy, witnessing a discovery a week during the early stages of the discipline; and theoretical physicists interpreting these results, refining their theories and suggesting new studies, as well as providing insight into the inner working of our universe. Months of delay in communication were out of the question: enter grey literature.
With the onset of use of the internet, the process continued electronically. And the community launched its own tools to manage this literature: in 1991, even before the web was invented, Paul Ginsparg, at the Los Alamos laboratories, launched arXiv, the first physics preprint repository[iii]. SPIRES[iv], the SLAC e-catalogue of preprints, which had existed since 1974, was able to benefit from the invention of the web, and become the first web server in the U.S. - it rapidly became the first HEP information system in the world. In 1993, CDS[v], the CERN Document Server, was launched.
Even if publication in journals is essential for the career evolution of HEP scholars, as for scholars in any other field, this is almost decoupled from the effective information transfer between scientists. Preprints remain today the main communication channel; "only" the discovery and submission processes have changed. Authors still submit a paper to arXiv before submitting to a journal: the date of submission of the paper to arXiv carries more importance to them than the "received" stamp on the subsequent journal article. Even in the era of electronic journals, therefore, grey literature fully retains its importance in the discipline. Instead of being phased out together with the paper era, grey literature is actually growing in importance, reflecting the evolution of modern scholarly communication beyond the printed word. Fifty years ago, scientists communicated by writing articles, of which the preprint versions made up the bulk of grey literature in the field. Nowadays, scientists go to conferences and their slides, generated in digital form, constitute a new form of grey literature that other scientists want to access immediately, and often to quote in their subsequent publications, without having to wait for a conference paper to be written and submitted as a preprint to a repository somewhere.

Grey literature has still today a paramount importance in scientific communication in HEP. Its discovery plays a pivotal role in the scientific workflow of HEP scholars. It is very interesting to understand how the community discovers and accesses grey literature, and whether current information services fulfil the information needs of HEP scholars and their projected evolutions in the new technological milieu. This is the aim of the survey jointly conducted in May 2007 several High Energy Physics institutes, which is discussed in this article.

The survey, methods and answers overview

From April 30 to June 11 2007, a survey was jointly conducted by several High Energy Physics institutes: CERN[vi] (European Organization for Nuclear Research), SLAC[vii] (Stanford Linear Accelerator Center), DESY[viii] (German Electron Synchrotron) and Fermilab[ix] (Fermi National Accelerator Laboratory). The aims of the survey were to understand the users' perceptions of current HEP information systems, to assess user requirements and preferences, and to define future needs.

This survey was promoted and distributed within the HEP community by e-mailing members of major experimental collaborations, users, and authors of papers at large laboratories, and by putting links on several of the websites under discussion, community-based as well as commercial. Publishers of popular journals in the field added links on their home pages and, in one case, e-mailed all registered authors. The survey included 20 questions, with both check-boxes and free-text answers possible; there were four optional questions in addition.

During the survey period, 2110 answers were received. It is estimated (by counting the number of distinct authors appearing on papers in SPIRES-HEP published in this decade) that the particle physics community represents some 20,000 people. Therefore the answers received represent over 10% of the community. This high participation is further enhanced by the fact that 90% of the respondents answered at least one free-text question, and 78% of the respondents answered at least one additional optional question. Over 50% of the respondents wanted to be contacted via e-mail to be informed of the results of the survey.

Furthermore, the answers came from more than 30 countries (Figure 1). This distribution is similar to that of the origin of published articles in HEP[x] which implies that the entire community is well represented by the answers. Of particular interest is the experience of users: are they new to the field, or have they been using these systems for a long time? 86% of the respondents have been using a HEP information system for more than 6 years, which corresponds to the expected age profile of the community, however, there is interesting information revealed by the behaviour of the younger physicists, as discussed later. Familiarity with the systems is demonstrated by the fact that 82% of the respondents use HEP information systems several times per week.

Given the volume of responses in comparison with the size of the community, and the interest shown by the participants, it becomes evident that the survey reached a substantial fraction of the community providing a real snapshot of its practices in the search and access of information and its information needs.

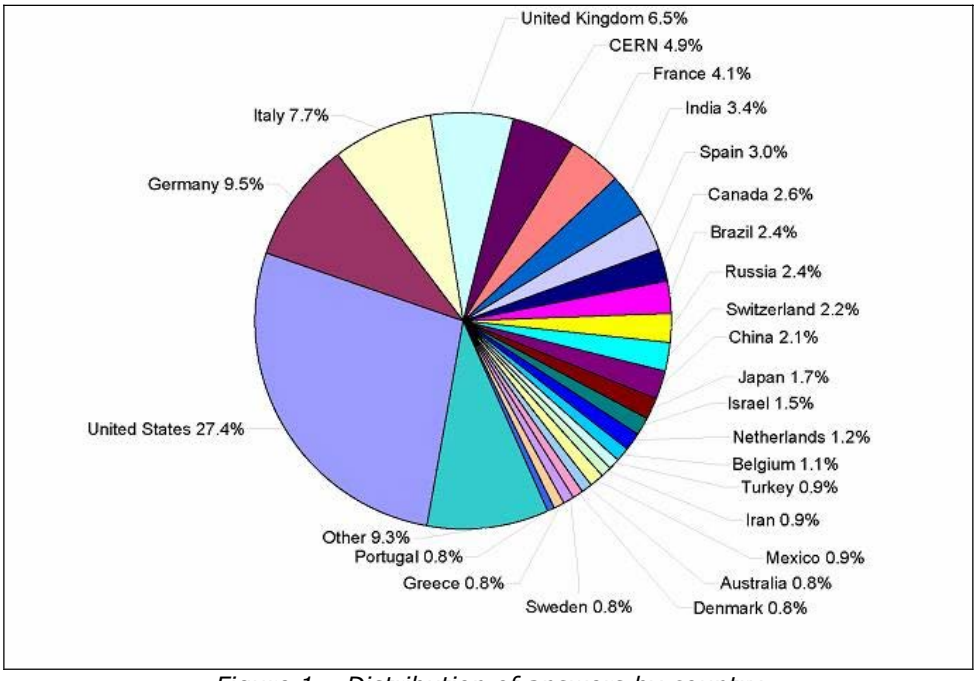


Figure 1 – Distribution of answers by country

The domination of community-based systems

The first and foremost question of the questionnaire was: 'Which information system do you use the most?' The results are presented in figure 2: 91.4% of the respondents prefer community-based services. This is not a surprising result since most of the community-based systems were created to meet the needs of HEP scholars, and they tend to be tailored specially to the evolving needs of the HEP community. These systems have been user-driven for decades. Google and Google scholar are used as the system of preference by 9% of the respondents. On the other hand, a mere 0.1% of users prefer commercial systems (commercial databases and publisher journal platforms).

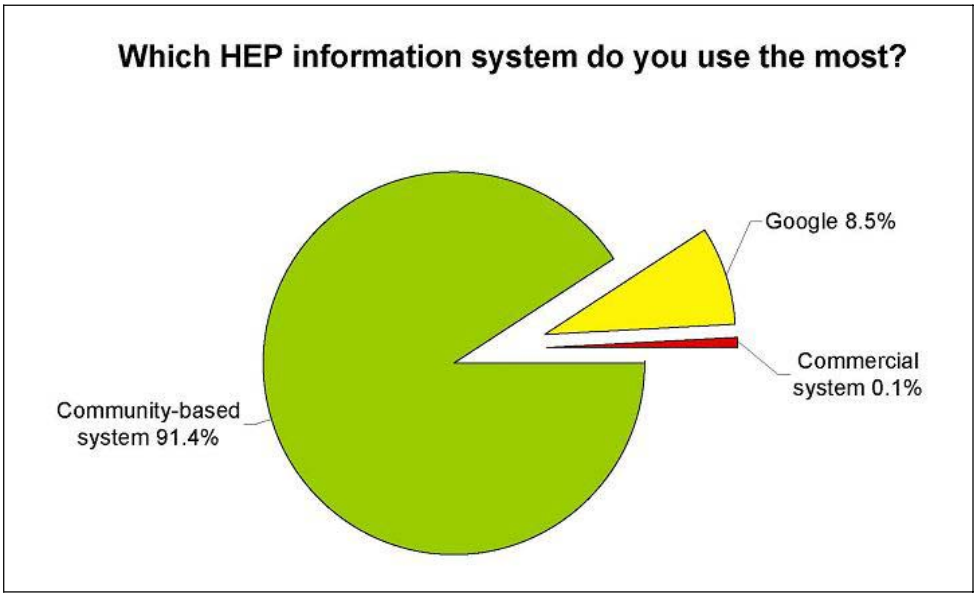


Figure 2: Scientific information systems favoured by HEP scholars.

Respondents were also asked which system they used the most when looking for preprints or published articles in different search situations (Figure 3). As one might expect, when searching for preprints, the community-based services are slightly more preferred than in the general case. The slice for commercial services run by publishers is only larger than one percent when a search for articles is concerned, and even then this amounts to only 3% of the total. Thus the overall landscape does not change substantially between preprint and article searching. This is extremely relevant in the interplay between grey literature and "conventional" literature in the field: when HEP scholars need to access "conventional" literature they still use the systems which were initially conceived to index and curate the grey literature in the field! Solutions invented for grey literature are therefore main-stream in this community.

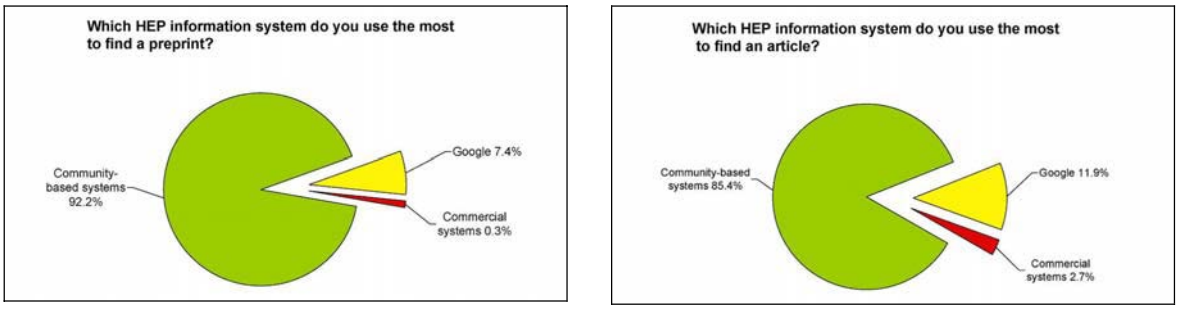


Figure 3 - Preferred information systems used by HEP scholars to find a preprint (left) or an article (right).

When respondents were asked about the systems used to look for theses (an example of "greyer" literature), the share of Google increases significantly, with Google and Google Scholar accounting for about one-third of usage. This emphasizes that community-based systems do not yet cover the complete scientific information needs of the field in the particularly important area of grey literature represented by theses. Community-based repositories can collect most HEP literature because it is, for the most part, on arXiv and in a few major journals. But it is a more labour intensive process to collect conference proceedings and theses, and so other automated search engines work relatively better here. In every other case, the field-specific utility provided by the community-based, tailored services overwhelms the often broader more expensive services offered by commercial providers. This conclusion has to inform the future development of the community-based services towards an investment in the harvesting and

curation of theses.
The correlation between the age of the respondents and their favoured systems presents an interesting trend, as shown in figure 4: the usage of Google tends to increase as the age of the respondents decreases. This is a reflection of the integration into the HEP community of individual scholars belonging to an age group which has been exposed to internet search before their academic career, as opposed to scholars who first used internet search engines during their professional activity. With the exception of theses, discussed above, it must be noted that Google might actually be only a gateway to other sources, as it indexes material from arXiv, SPIRES and CDS. Effectively, those using Google ultimately use some of the other information services to really access the document.

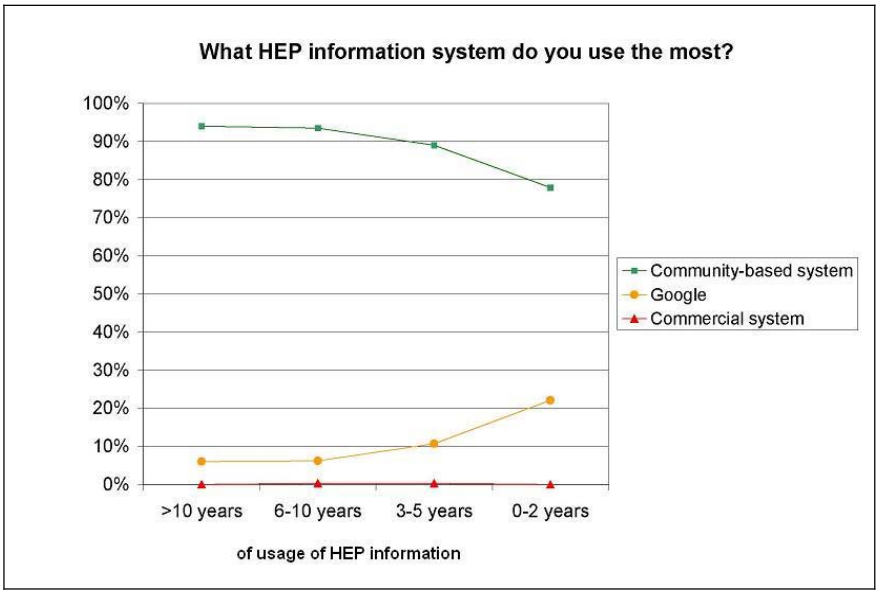


Figure 4: Comparison of the preferred system with the number of years a scholar has been active.

In summary, community-based systems largely dominate the landscape, even if Google takes a non-negligible part. The choice of information systems is linked to needs, and varies according to what scholars are looking for, each information system having strengths and weaknesses, depending on the features they have developed.

The current and future needs
The central part of the survey asked the users to prioritize the features of a scientific information system most important for their research workflow. The results are summarized in Figure 5 and show that 'Access to full text' comes as a first priority. It is therefore not surprising that community-based systems have such an important place, as they all aim to give the easiest and fastest access to the full text of documents. The same applies to the second feature: "Depth of coverage", indeed HEP repositories give very close to full preprint coverage. This applies both to arXiv.org, which in the last decade received submission of virtually all preprints relevant to the field, and to SPIRES, whose staff have painstakingly curated a complete database of all literature relevant to HEP scholars. These systems are field-specific, with searches returning close to zero noise, and therefore obviously fulfil the demand for: 'Quality of content'.
The three winning features therefore correspond exactly to the design requirements of community-based systems; this explains why these systems have such an important place in the landscape. These strong user requirements will continue to inform the evolution of these systems.

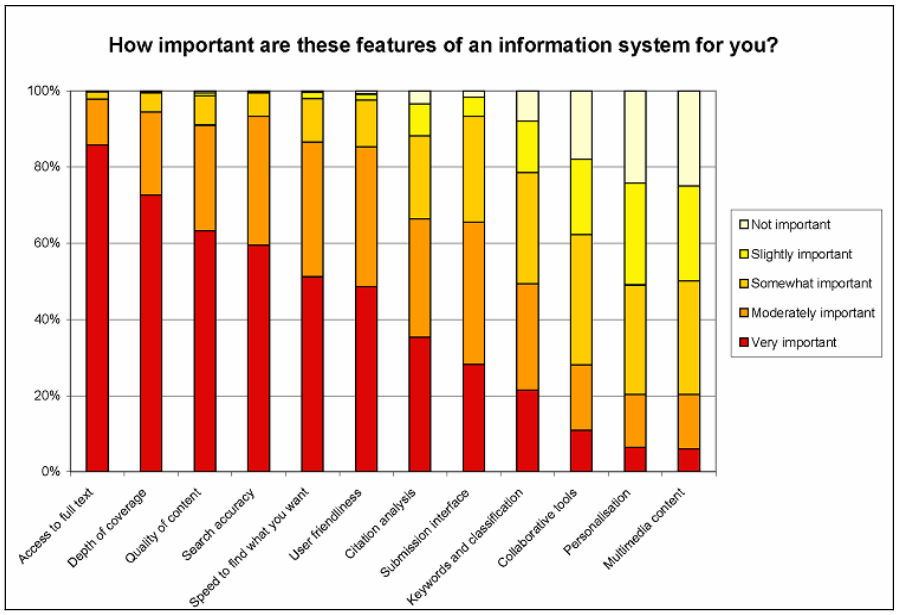


Figure 5 – Most important features of an information system

A crucial question in the survey was whether users expect some changes in their information needs in the next five years: 75% answered positively. Times are changing and the users are well aware of this, posing a challenging environment for the developer of scientific information systems.
Finally, the survey inquired about the future needs of the users. The analysis of these answers is still ongoing and only an overview is presented in Figure 6. Most of the projected needs are linked to access to full text and deeper content: "linked presentations of all instances of a result", "access to data and tables". According to free-text answers, access to more grey literature is a very strong need; some users emphasized the importance of having access to conference slides, theses or laboratory reports. This is again proof of the growing importance of grey literature for the immediacy of communication in the discipline in the new technological environment and the strong need of HEP scholars to easily access these information resources.
It is interesting to remark that centralization of information access comes in second place: 70% of the respondents think that it is somewhat important. This shows that users would like to use only one information gateway. This need explains to some extent the growing popularity of Google, where people go for a 'primary search' and are redirected to the right information system. Google is, in this case, a metasearch engine: a simple gateway to other tools.
As a very first conclusion, we can say that the future needs imagined by users correspond essentially to a wish to have easier and wider access to content and data. But inspired by the onset in our daily communication channels of the web 2.0 paradigm, other needs appear, such as 'recommendation of article' (almost 50% of the respondents think it is somewhat important).

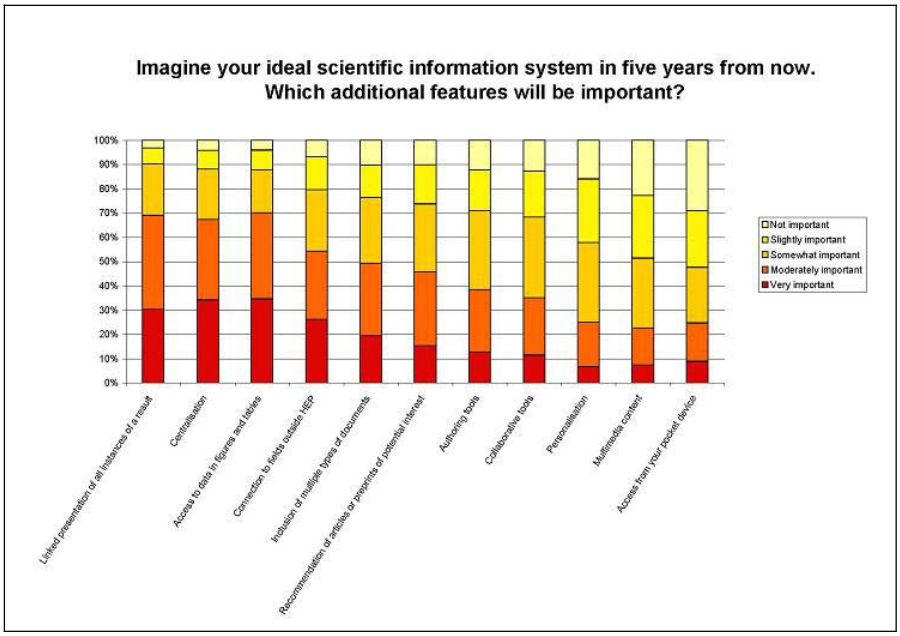


Figure 7 – The features of an ideal information system

Conclusion

The overwhelming number of answers to this survey was almost unexpected, reaching 10% of the community. This shows an incredible level of interest and involvement of the community in informing the evolution of tools which are central in their research workflow. HEP scholars are very discerning in their information needs and know what they want. This article presents some preliminary and partial results of the survey. The analysis of the responses is still ongoing and more complete analyses will soon appear. However, the current status of information search in the field and the future needs of the community are very clear.

Community-based services are overwhelmingly dominant over Google and commercial services. In other words, the HEP community is able to fulfil its own needs. Users said that their primary concern is access to full text. However, other features, mainly linked to the evolution of grey literature in the new technological environment, will be important in the coming years. Therefore, changes have to be expected. To continue writing pages in their multi-decade success story, community-based systems have to evolve with the needs of the users, meeting the Google-generation's wish for centralization of information. Community-based systems could meet this requirement by uniting their efforts and trying to construct a unique information platform that would cover all the needs of the users, present and future. This new system should continue to be designed around "old" and "new" forms of grey literature, enabling on one side, new usages of electronic preprints, including exploitation of text-mining applications, and on the other, an exploration of "new" forms of grey literature, such as, but not limited to, slides presented at conferences. Finally, these systems should brace themselves for the next generation of users, who will expect web 2.0 features in all their information systems.

References

[i] We use the term "information service" to describe this broad class of services that provide researchers with information or access to information that they need for research. This concerns only information services relevant for HEP community

[ii]Luisella Goldschmidt-Clermont, Communication Patterns in High-Energy Physics, published in *High Energy Physics Libraries Webzine*, 6 (2002). <http://library.cern.ch/HEPLW/6/papers/1/>

[iii] Paul Ginsparg, First steps towards electronic research communication, published in *Computers in Physics* 8 (1994) 390.

[iv] SPIRES-HEP database has been run by the Stanford Linear Accelerator Center <http://www.slac.stanford.edu/spires/>, with Fermilab and DESY. The history of SPIRES, including how it came to be the first web server on US soil is recounted in L. Addis, 2002 <http://www.slac.stanford.edu/spires/papers/history.html>

[v]CDS is run by CERN <http://cdsweb.cern.ch>

[vi] <http://www.cern.ch>

[vii] <http://www.slac.stanford.edu/>

[viii] <http://www.desy.de/>

[ix] <http://www.fnal.gov/>

[x] Quantitative Study of the Geographical Distribution of the Authorship of High-Energy Physics Journals, Mele S.; Lindqvist, C-M; Krause, J, July 2007**.** CERN-OPEN-2007-014

Virtual Reality and Establishing a Presence in Second Life: New Forms of Grey Literature?

Kristine Ferry, Julia Gelfand, Dana Peterman, and Holly Tomren
University of California at Irvine, United States

Abstract
The virtual reality program Second Life poses new challenges to its more than 3 million users that include economic interactions, methods of communication, and documentation. This paper assesses the range of activities taking place in Second Life, proposes how Second Life may be viewed as a form of emergent grey literature, discusses the tools used to discover, document, and preserve it.

Just five years ago, we would not have such a presentation as I am to share with you today. Virtual Reality, much less, Second Life was just in the embryonic stages of development, making today, just beyond toddler years. The early stages we have seen pioneered have brought to the forefront some new roles that have now become mainstream in education and learning and that is what the focus is today. Virtual reality is defined as a technology which allows a user to interact with a computer-simulated environment, be it a real or imagined one, and can be similar to the real world or differ significantly from reality. Common methods of engaging in virtual reality is via computer games and the technical limitations of creating certain kinds of environments is shrinking daily as faster and more advanced computer processors, improved imaging and data communication technologies become more ubiquitous and affordable in each new release.

Early experiences in virtual reality have been captured and developed in several now classical books, including those by Howard Rheingold (*Virtual Reality*, 1991; *Virtual Community*, 1993) who demystified the subject and made it part of the average college student's lexicon. Many different applications of computer graphics depicted over the last decade led to advancements in audio and sound amplification and thus a higher quality simulation was possible. As entertainment and education entered the world of virtual reality in a profound way, the futurist Mychilo Stephenson Cline (2004) offers how cognitive and behavior scientists were exploring how virtual reality leads to a number of important changes in human life and activity by being more integrated into daily life with patterns and techniques to influence human behavior, interpersonal communication and cognition.

Science fiction, film, music videos, computer games, art, marketing and promotions each are representative examples how experiments in 3D opened up new creative opportunities. By 2003, when Philip Rosedale started Linden Research, now known as Linden Labs to launch Second Life (http://secondlife.com/) virtual worlds were well on the generation of youth's radar and it was the new environment in which people could interact, play, engage in business, study and learn, and otherwise do business, and it was this concept "that Second Life would be like reality, but kinder and gentler" (Maney, 2007). By populating it with residents, genuine people who have names and powers and known as avatars, a new civilization was born that could be of either sex, customized to produce a wide variety of humanoid and other forms in a variety of costumes to resemble both real and imaginary life (Second Life avatars and their real life, 2007).

Second Life has its own method of commerce trading in Linden dollars and the environment is divided into islands forming different real estate parcels that can be bought and sold, resembling the rise and tumble of real life real estate transactions in this volatile age of disputes over land use. Residents must develop skills to build virtual objects or anything that they want on or to furnish their parcel and a 3D modeling tool helps do that, or now in a more mature virtual world, one can buy services and products already produced but the resident avatars are responsible for creating their environments.

Commitment to SL in Business

Brick and mortar businesses have been quick to take advantage of Second Life as a virtual environment to conduct both their internal and external business. Companies such as IBM and Reuters are using this virtual world to conduct recruiting for a variety of positions (Corporations use Second Life as recruitment tool**,** 2007). Some in human resources even believe that virtual worlds like Second Life can't be beat as a non-discriminatory method of interacting with potential employees and clients (Krell, 2007). In addition, companies are using Second Life as a tool to conduct meetings among geographically dispersed workers. American Apparel was one of the first companies to set up shop in Second Life and begin selling its virtual merchandise. Other companies are now offering products and services for sale from clothes to law advice. For profit businesses aren't the only ones getting involved. Charitable organizations such as the American Cancer Society are also actively involved. The American Cancer Society hopes to raise $75,000 in 2007 (Murray, 2007). In fact magazines and journals devoted to those in the non-profit world are now offering advice on how to make the most of this new frontier underlining that it is not one that should be ignored (Rigby, 2007).

That being said, much work needs to be done in studying how people and organizations are behaving in Second Life. Are they following the same social norms as they do in the real world? Is there anarchy or is everyone living together peacefully? Is anyone watching out for people or is everyone on their own? There are some articles beginning to appear that discuss how behavior in the virtual world should or should not be similar to the real world. These issues and more are being explored. Linden Labs has created a document called its Community Standards that outlines six behaviors that will not be permitted within Second Life: intolerance, harassment, assault, disclosure, indecency and disturbing the peace. How can businesses or people conducting their own business make sure they are following such standards? One example of a corporation doing what they can is IBM who has issued etiquette guidelines for its employees to use when in Second Life (McConnon, 2007). However residents have to be responsible for their own actions and safety to some extent just as they do in the real world. Harm could come in the form of any of the behaviors Linden Labs outlaws. Most noticeable so far has been fraud. An article in the *Illinois Business Law Journal* describes how one of the banks within second life is most likely a Ponzi scheme which pays off original investors with money contributed by new investors. In such schemes the new investors don't end up getting paid. How can people or should we say, avatars, protect themselves? There is a mechanism in Second Life that allows residents to report abuse and abusers can actually be banished from this virtual world but it is unclear if they could be punished legally in the real world. Legal disputes are starting to show up pertaining to Second Life but taking place in real courts including cases involving land disputes and copyright infringements. For example, people are creating objects that could infringe on trademark law by using famous logos on virtual versions of their favorite stuff (Ward, 2007). Lawyers are preparing for a possible increase in this trend. An article recently appeared in the *Brooklyn Law Review* discussing the legal implications of participating in Second Life (Chin, 2007) showing that they think this phenomenon is not going away. Britain's Fraud Advisory Panel has gone so far as to advise the government to put real-life financial regulations into place saying Second Life is "a parallel universe with almost no external rule of law, no enforced banking regulations or compliance, no policing and no governmental oversight" (Leapman, 2007).

Geography is important in Second Life. Address, adjacencies and location play vital roles as navigation is achieved by different methods, for an avatar can walk, run, and fly to different points. These methods of transportation help define the organizational life and become one of the ways that business and industry, government, public policy and educational life have each flourished in Second Life. Today, many different forms of learning and teaching take place in Second Life and non-profit organizations have a strong presence. The full range of education can be tracked ranging from juvenile or Second Life for Kids where permissions are required for residents under age 16, and among the more active educators appears to be librarians. There is an InfoIsland region where many libraries have invested and established themselves, and still many others have a presence elsewhere. The directories that try and monitor the range of institutional property holders are never accurate and the best way to find this information is to search within Second Life.

Grey literature, also known as fugitive, hidden, invisible content, in the deep web or hard to identify, find, acquire and access, has some parallels in learning how to determine what is actually transpiring in these examples of virtual worlds. As they proliferate and more organizations and individuals are identifying a presence there, it becomes increasingly curious and of interest to assess the level of activity and by whom, when and where in environments such as Second Life. Some analogies of other new forms of electronic communication and new media that require and now offer indexing and finding tools are blogs, wikis, podcasts, among others, forming new frameworks for reference resources. The Web 2.0 tools allow one to better organize and describe elements at a more finite or macro level and permute, expand and cascade relationships of terminology and vocabulary. Metadata, once the domain of indexing and the assignment of descriptors have expanded into user or reader initiated social tagging and tag clouding to personalize content an individual has found, wants to save, recall, and potentially repurpose or reuse. Bibliographic management software packages were an attempt to organize traditional literature by creating databases of citations and entries from multiple sources that could be saved, recalled and cited in various ways in future writing needs. With bookmarking, music tracks or compositions and image finding tools such as del.icio.us, Last.fm or Flickr tagging has come to be associated with pieces of relevant information that is just one component of a bigger enterprise, such as a photograph, map location, blog entry or video clip with customized keyword-based classification methods. These resources are more personal, using the terminology the user adopts to describe content rather than a dictionary that may be foreign or less familiar and thus intimidating, and contribute to the new suite of folksonomies and user-based tagging that is increasingly common.

Applying this form of categorization to Second Life to determine if something there may be of interest has great potential and utility. If classes, webinars, tours, business transactions, and information sharing take place in Second Life, how does one find them?

We cannot know whether Second Life, or any virtual reality for that matter, is gray literature and not ephemra, because as technology develops and our expectations of document longevity and legitimacy

shift, the differences between them become ever murkier. Virtual reality is a wonderful expression of postmodernism, even though we know that postmodernism cannot exist outside of historical context – a conundrum. Virtual reality in Second Life, as elsewhere, then, is a kind of folded mirror in a funhouse. The meaning of Second Life as both a virtual reality and a grey literature relies on its replication, modeling or collapsing of the original experience into recognizable symbols and signs over time. What signs, symbols, and documents can represent something that so closely resembles our lives, our thoughts and impressions in such a venue without being the thing itself?

So, to answer that question within the context of Second Life, like archaeologists we dig up conventional and unconventional markers that constitute its existence as a form of literature. As a consequence of its literary infancy, we find ourselves having to locate Second Life's authenticity in older forms – we relied on books to explain the World Wide Web, now we rely on the Web to explain and validate Second Life and its activities. Looking at our "dig," we can infer why Second Life is in this liminal state. Grey literature, or any medium for that matter, requires at least four interlocking and overlapping components: a community of individuals; a sustained period of learning that generates outside referents; a certain amount of time for that medium to exist; and a "presence" that has both longevity and persistence.

Outside of Second Life, grey literature is created and maintained by community members who are early adopters, people who live and perform multiple literacies and who are living in a liminal state. These early adopters provide the impetus for sustained growth and help others to create negotiated meanings. They are the ones who straddle both worlds by leaving behind material evidence of the newer form in other media and who use organizational schema to describe it. Not surprisingly, we find librarians and educators, adapting and sustaining the medium.

There can't be just a few people though, the community must grow to some critical mass that generates materials and tools to make functioning in that new environment more accessible and entertaining to develop social networks that have meaning. So, Linden Labs offers with easily customizable avatars that take on characteristics and abilities that are more interesting than that of our first lives. We can fly. We can teleport. We can make objects that interact. We can destroy. We can express many stories and live narratives by being several selves, with our multiple avatars represented not just by text, but by radically different appearances and abilities.

Most significantly for a community, however, we can interact with others. More (or less) importantly to history, we can bring those experiences back into other venues such as books, blogs, and other social networks to name a few. In the midst of this ready-made production, community members with greater skills that allow them to build objects and experiences scaffold and enrich the Second Life experience for newer initiates who, in turn, continue to reshape the community. This liminality is the birth of desire, and of replication.

This desire to replicate comes at a high cost, though. It takes a long time to acclimate to the use of Second Life, and it requires sustained learning to contribute to a social network of this kind. It is very similar to drawing easels and typewriters, with no quickly discernable utility, but with a lot of creative potential. There are things to learn about the environment; such as building objects and communicating through notecards, text, gesture, and sound; such as the mutable identities that can be assumed. That information has been necessarily replicated and changed outside of the new context. Linden Labs knew this when it wrote a book and created its blog, and others quickly enforced that replication by following suit with websites, YouTube productions and Machinima (aka virtual filming). The detritus of this learning has longevity and contributes to what we think of as grey literature.

The popularity and the effort put into creation must be sustained over time to develop meanings. Second Life has been in existence since 2003, but has this been long enough for it to be literature of any kind? It has and it has not if we begin to consider recent forms and mediums. At this juncture, most of us can barely remember the existence of a life without the World Wide Web, established a decade before Second Life. But, how many of us remember with great clarity the browsers of old, like Mosaic or Opera? If we are to believe that technology progresses in a geometric fashion, and that Second Life is sufficiently robust to continue to dominate this platform, then it will survive in some form to translate into other forms of virtual reality. So, Second Life has a limited amount of market dominance time to create persistent presence.

Presence is embodied within Second Life in conventional and unconventional manners. Signs are posted on land; "notecards" are obtained from objects – both conventional texts. Unconventionally, however, the experience of a tour, the layout of land, and the description of the self, are also presences. Even events could be considered a presence. The documentation of all of these rests both within the object and outside of it. How such objects and experiences are found create the "greyness" of literature and it

shares attributes with ephemra. However, we can see that activity and commitment within Second Life, and other virtual realities, is robust based on the economic investments that are being made.

Second Life's search features are currently limited in scope. Only certain kinds of data are currently indexed and searchable within Second Life, and the user-supplied metadata that informs these searches is often incomplete and unreliable.

In the current version of Second Life Search, a user can search for places, events, people, and classified ads. Objects are not currently searchable within Second Life; however, to compensate for this, some third party companies such as Electric Sheep (http://www.electricsheepcompany.com/) index objects, but only for-sale objects on "public" land parcels.

Searching in Second Life is soon to see some major changes and improvements. Linden Labs announced on October 19, 2007 (http://blog.secondlife.com/2007/10/19/new-search-currently-under-development/) that a new in-world search mechanism is being developed and may be available in late 2007. Linden Labs has built a system that extracts Second Life data to HTML Web pages and uploads these Web pages to a Web server, where they will be indexed using Google Search Appliances, an off-the-shelf product which Google makes available for purchase. This relationship remains curious, as Second Life Linden Labs is not in a working partnership with Google. The new search mechanism promises to include land parcels, resident profiles, groups, Second Life wiki articles, events, and some objects present on "public" parcels. This will be the first time objects will be searchable within Second Life's search function. Users will have the ability to specify whether certain avatar profile information and objects will be searchable. Because of the search functionality that comes with using a Google product, more advanced search queries will be possible, based on Google search operators. In addition to in-world searching, anyone with a Web browser will also be able to conduct the same searches on the Second Life Web site, and search results may also be indexed by external search engines such as Google and Yahoo.

Places in Second Life are currently keyword searchable by two indexed user-supplied metadata fields: Name and Description. However, places are only searchable if the land owner selects "show in search," which costs 30 Linden dollars per week (approximately $0.10). Therefore, places in Second Life are by default not searchable; land owners must opt in to Place Search by paying a fee. Parcels in which the land owner has opted to pay for search are referred to as "public" parcels. The land owner must supply a parcel name, but a free text description (up to 120 characters) is optional. Land owners could certainly increase their land's findability by enriching the data in the Description field, but some users leave this area blank. The new search will not significantly change Place Search.

User profiles are currently only searchable by avatar name. In the new Second Life search, additional descriptive information in user profiles will also be keyword searchable, but only if users opt in to allow their profile information to be included. For users who opt in and also provide detailed profile information, this potentially creates increased social networking opportunities, where users may be able to find each other based on common interests and activities.

Second Life "objects" are currently not searchable by the Second Life search function. The new search mechanism will include some objects, but not all. Objects that are listed as "for sale" and that reside on "public" (paid to be searchable) parcels will be searchable by default. Residents will be able to change the settings for both for-sale and not-for-sale objects so that they will show or not show in search, but objects will only show in search if the land itself is also set to show in search (by paying to be listed). Objects are even more problematic than land parcels in terms of user-supplied metadata. Similar to land parcels, two fields are available to describe objects: Name and Description. By default, the name of an object is "Object" and the description is blank. Many users do not bother to change the default text or supply additional descriptive metadata for the objects they create. When the new search is launched and a large number of objects simply named "Object" are displayed in search results, perhaps some users will pay more attention to the data in these fields to increase relevant search results.

Notecards are a major category of content contained in Second Life and are not now searchable, nor will this functionality be included in the new Second Life search. Notecards themselves have Name and Description metadata fields, for which, similar to objects, many users do not fill in or change the default text. Besides these fields, notecards are a rich source of textual information in the body of the notecard itself. Some libraries having a presence in Second Life have created "books" in Second Life, in which the text of the book is made up by notecards. Many Second Life sites offer informational notecards of various sorts, but this content is not full text searchable in any way.

In addition to problems of searching within Second Life, Second Life content is not truly preservable or accessible outside of Second Life. Avatars, objects, notecards, and landscapes cannot be exported out of Second Life in their native form (though Linden Labs is interested in the concept of "universal avatars" that could be used in other Web and gaming applications -

http://www.virtualworldsnews.com/2007/10/ibm-and-linde-1.html). If an event takes place in Second Life, several means could be employed to attempt to record it: taking screenshots and posting them on sites such as Flickr; recording the audio content, if any, and offering it online as a sound file; using video recording software to create a digital video or "Machinima" of the event, and posting the video content on sites such as YouTube; and saving the text of a chat transcript and pasting it to a document file, Web page, or blog post. In some ways, this attempt to make Second Life experiences available outside of Second Life may be more retrievable outside Second Life than within it, due in part to user tagging/folksonomies available on Web 2.0 sites and due to the superiority of most Internet search engines to Second Life's built in search capabilities. However, these attempts to replicate Second Life experience outside of Second Life are scattered on the Web and users' personal computers and cannot fully reconstitute the Second Life experience.

The problems associated with searching within and preserving Second Life content may be partially explained by a model that is geared towards commerce rather than information seeking and organization. The assumption implicit in Second Life's built in search function, whether the current search or the soon-to-be-unveiled new and improved search, seems to be that the primary things for which users are searching are things to buy and sell. However, there are a great many Second Life users creating content in Second Life that has an educational focus and exists for reasons other than being bought and sold. This informational content is lost in the current model.

Despite the pending improvements in Second Life search, several key issues ensure that Second Life content will remain in the realm of grey literature:

- User-supplied metadata is incomplete and of inconsistent quality, which decreases searchability ("garbage in, garbage out").
- Relative to the content available in Second Life, very few types of information are searchable; most notable may be the lack of full text searching on notecard content.
- Much Second Life content, including land parcels and objects, is not searchable by default; residents must pay a fee, however nominal, for this privilege.
- Second Life content and experiences cannot truly exist out of Second Life; attempts to replicate parts of it are scattered throughout the Internet.

Conclusion

We are uncertain that Second Life creates or is a grey literature, but we know that it shares attributes with the standard definition of grey literature (Luxemburg). The markers of a new form: communities, education, longevity, presence, exist, but not necessarily to the extent that we would expect.

Second Life's documentation and bibliographic control continues to shift and is somewhat hampered by its early commercial origins. Because it is explicitly, but not entirely, a commercial venture it has potential. Commercial ventures evaporate without the support of noncommercial motivation. The things that most clearly delineate our understanding of Second Life are comparatively more "bibliographically" controlled than the thing itself. Monopolistic search engines like Google may overwhelm or ensure its existence. Its popularity and diffusion rate remains in flux.

Second Life's incorporation into real life economics and the growth of a purely virtual marketplace ensures that some mark will be made, but not necessarily anything outside the purchase of a few McMansions. We remain uncertain of its validity as a form of literature and unclear that its documentation or preservation is the responsibility of any particular entities, however it will retain a sense of space.

References

Chin, B. (2007). Regulating your Second Life: Defamation in virtual worlds. *Brooklyn Law Review,* Retrieved November 27, 2007 from Westlaw Campus database.

Cline, M. S. (2004). *The future of virtual reality*. Unpublished master's thesis, San Jose State University, California.

Corporations use Second Life as recruitment tool. (2007). [Electronic version]. *FinancialWire,* Retrieved November 27 2007 from Factiva database.

Krell, E. (2007). HR challenges in virtual worlds. *HR Magazine, 52*(11) Retrieved November 27, 2007 from Business Source Premier database.

Leapman, B. (2007, May 13). Fears that virtual world will become a haven for terrorists Second Life game's scope for money laundering and tax evasion is all too real, say fraud advisers. [Electronic version]. *The Sunday Telegraph,* Retrieved November 27 2007 from Factiva database.

Maney, K. (2007, February 2). The King of Alter Egos is surprisingly humble guy: Creator of Second Life's goal? Just to reach people. *USA Today*, p. 1B.

McConnon, A. (2007, August 13). And now, a virtual dress code. [Electronic version]. *BusinessWeek,* Retrieved November 27 2007 from Factiva database.

Murray, S. (2007, May 25). Charities take their events online. [Electronic version]. *Financial Times,* Retrieved November 27, 2007 from Factiva database.

Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison-Welsey.

Rheingold, H. (1991). *Virtual reality.* New York: Simon & Schuster.

Rigby, E. (2007). At work: Fundraising - how to ... carry out fundraising through online communities. *Third Sector,* Retrieved November 27, 2007 from Factiva database.

Second Life avatars and their real life. (2007, June 17). Web posting to Web Urbanist to http://weburbanist.com/2007/06/17/top-3-look-alike-avatars-and-people-from-second-life-to-real-life/

Ward, S. F. (2007, March). Fantasy life, real law. *ABA Journal, 93*(3), 42-47.

Finding the Grey in the Blue: Transparency and Disclosure in Policing

Paul Sturges and Louise Cooke
Loughborough University, United Kingdom

Abstract
Police services have traditionally valued the ability to work without ongoing public scrutiny of their investigations and operations. They can very reasonably cite the need to avoid alerting criminals to police activities that might result in their arrest and charging with offences, the need to protect police and witness safety, and the frequent need to act swiftly and decisively without obtaining special approval from relevant authorities or endorsement from public opinion. This necessary lack of disclosure concerning many police operations has often extended into a general lack of transparency regarding police activities and expenditures, to the extent that, in many countries, the police services are regarded as unaccountable and unconcerned with how public opinion perceives them. In such a climate, police corruption and arbitrary exercise of police power flourishes. This paper addresses the creation of a policing environment radically different from this through the introduction of transparency into policing in the UK and the consequent revelation of layers of grey documentation and data. The paper makes use of official documentation and case studies of selected British police forces to show how the culture of policing is being changed. The principles of open government, scrutiny, and disclosure with a view to establishing accountability, are in the process of becoming institutionalised in the UK right across government, local government, other 'public authorities' and the business and non-governmental organisation (NGO) sectors. The UK Human Rights Act 1998 sets the context, and a legal framework for this transparency is provided by the Freedom of Information Act 2000 and, to some extent, the Public Interest Disclosure Act 1998. The press and civil society are consistently using these mechanisms to call those with political and economic power to account. It has become apparent, even in sectors formerly as concerned with avoiding openness as the police service, that pro-active disclosure is the best way to meet public expectations. Police services now respond as a matter of course to freedom of information requests, organise a range of meetings to provide information and answer questions (from local officers' meetings with community groups through to major budget consultative meetings with citizens' panels), and participate in public and semi-public enquiries into aspects of the success or failure of police programmes and operations. The case studies in this paper will explore the opinions of key players in this process and draw attention to the grey information that is becoming available as a consequence.

INTRODUCTION

British policing is now very different from the period before 2000, and this is largely an information phenomenon. Despite the fictional image of days and nights on patrol, fights and car chases, and cases solved by the brilliant insights of individuals, policing has always been essentially an information handling profession. The last few years have merely served to bring information aspects to the fore, through the strengthening systems and, most significantly, the introduction of high levels of transparency. The passing of the UK Freedom of Information Act in 2000 may, at first, seem to be the driving force behind this, but there are other influences and pressures that are arguably more significant. This paper concentrates on transparency and disclosure in policing to demonstrate the ways in which publications and semi-published, broadcast and electronic information arising from police work have been made available to the British public in ways which have genuinely revealed the grey resources that can inform people about activity in the blue-uniformed ranks of the police. All of this is comparatively new because in the past the need for not subjecting ongoing police investigations, and sometimes aspects of the judicial process, to public scrutiny has dominated the police information ethos.

In this paper we explore the shift from an ethos of concealment and confidentiality to one of transparency and disclosure mainly through the cases of two sample police forces. Their grey literature and web resources will be identified and described in the light of opinion and contextual information obtained through interviews with personnel from those forces. The paper is openly more descriptive than it is analytical. As a first step into this new and little investigated (at least as far as information science is concerned) information environment, it is worthwhile simply to describe and make some sort of initial assessment of the dimensions of the phenomenon. The content of this paper is largely drawn from observation and discussion of the activities of two police forces in the English Midlands: Derbyshire (1) and Leicestershire (2). The information materials and access obtained for these two case studies have been examined and placed in context through meetings with communications staff in both forces. The demography of the two counties makes them reasonably representative of the country (they are in the mid-lands in more senses than one) and the statistics relating to policing confirm them as worthwhile examples. Generalisations made from two cases are, of course, vulnerable to later more extensive and intensive research, but for the purposes of a preliminary essay we believe that these cases serve very well.

BACKGROUND: POLICING AS AN INFORMATION PROFESSION

The effective management and use of information has always been critical to crime prevention, reduction and investigation strategies. The 'paper burden' borne by police officers is heavy, even in these times of electronic communication (Luen and Al-Hawamdeh, 2001; Ericson and Haggarty, 1997). A recent study within a UK policing unit indicated that personnel regarded information overload as a significant barrier to effective internal communication within the unit (Syson, 2007). As a result of this avalanche of paperwork it is generally acknowledged that the police spend a relatively small proportion of their time dealing directly with crime (Ericson and Haggerty, 1997). Indeed, it has been suggested that direct involvement in crime work takes up as little as three per cent of the working time of patrol officers (Comrie and Kings, 1975; Ericson, 1982, cited in Ericson and Haggerty, 1997, p.20). Even where criminal investigation work is considered, Ericson found that detectives spent about half their time in the office, much of it on recording investigation activities rather than on conducting actual investigation work (Ericson, 1993). An example of this can be seen in the requirement in England and Wales for all interviews with suspects to be audio taped and then summarized on a standard form. The time devoted to this one specific requirement has been estimated to be equivalent to the full-time efforts of 1,400 police officers (Royal Commission on Criminal Justice, cited in Ericson, 1997, p.20).

At the local level, it must be recognised that the operational and regulatory environment in which police officers work is also dynamic and fast-changing, complex, sensitive and highly stressful. Such work frequently requires the assimilation, management and generation of vast quantities of information in a timely and effective manner. Thus, despite the contention of Brodeur and Dupont (2006) that informational work ('shuffling paper') is perceived as being of little value and carries no prestige within policing organisations, it is clear that policing is indeed an information-intensive operation.

However, whereas in the past much of this informational activity, and the records arising from it, were not open to public scrutiny, police forces in the UK are now operating in a climate where public bodies and members of the public demand a high level of transparency and information disclosure. High profile incidences of failure to pass on information between police forces and other agencies have increased the need to scrutinise the external communication patterns of policing organisations. Thus, for example, the terms of reference of the Bichard enquiry, investigating the circumstances relating to child murders in the village of Soham in Cambridgeshire, included an assessment of police information sharing with other agencies. This concluded that the lack of clear, understandable, national standards and guidance on the subject of information sharing and of record creation, retention, review and deletion, had contributed to the tragedy (Bichard, 2004). This recognition led to the drafting of a Code of Practice on the Management of Police Information (MOPI) (National Centre for Policing Excellence, 2005), together with guidelines for implementation of the Code (National Centre for Policing Excellence, 2006), which challenged the UK police service to move forward from a culture of information storage to one of information management and sharing. In addition to ensuring consistency between Forces in the way that information is managed, the Code of Practice aimed 'to ensure effective use of available information within and between individual forces and other agencies, and to provide fair treatment to members of the public' (MOPI, 1.1.4). The subsequent implementation guidelines give recognition to the value of information as a 'corporate resource' for the police service, and the principle that effective policing is dependent on the efficient management of information.

ORIGINS OF CHANGE

The emergence of police forces which are both more effective as handlers of information and more open to the scrutiny of outsiders than before is still a work in progress. However, from many (if not all) British police forces there is now a confident projection of themselves and their activities as a matter for sharing with a multiplicity of stakeholders. When asked the question as to the sources of this change, the response is sufficiently diverse and, indeed, imprecise as to suggest that the new openness arises from a broad cultural change across the sector, rather than a single driving force. In the past there was formal communication with the public driven by the legal requirements to report on quantitative and financial aspects of the work of each police force. Anything else seemed to be regarded as potentially confidential, or at least not meeting some unspecific test of a public 'need to know'. As we have already indicated, government priorities have shifted the emphasis towards disclosure and consultation. There is, for instance, a requirement to hold public consultative meetings. Senior appointments now tend to be made from officers who buy into an ethos of openness, and this has quite quickly influenced practice. Most specifically it has led to the creation of communication teams dedicated to involving and informing the public.

Legislation not specifically directed at police forces has also contributed to this fast-moving change: first of all the relevant clauses of the Human Rights Act, 1998, the contribution to this ethos of the Public Interest Disclosure Act, 1998, and most significantly the Freedom of Information Act 2000. Relevant policy initiatives include the *Modernising Government* agenda of the Blair government, which focussed on a

commitment to public sector reform, including making public sector bodies more accountable to their users (Parliament, 1999). This in turn led to a programme of police reform outlined in a White Paper *Building Communities, Beating Crime: A Better Police Service for the 21st Century*, which aimed to build 'a more responsive police service with neighbourhood policing at its heart' and to improve the way the police are held to account (Home Office, 2004). Specifically, with regard to information provision, the White Paper commented that

> 'Local people need to be clear who is responsible for what in terms of community safety. They need to understand how they as individuals, families and members of the community, can play a role in keeping their communities safe and reducing crime; how they can have a say in setting local priorities, and how well their local police are performing. This information needs to be available to every household and people should know what they can do…if local problems are not being tackled effectively.' (Ibid., p.10)

The increased capacity to communicate demanded by such measures has, in turn contributed to an increased ability to address priorities through the means of communication. Intense public interest in policing, and vociferous criticism of perceived failures and inadequacies, reflected in and sometimes fuelled by the media, have demanded a response. In some parts of the world this has undoubtedly led police forces to retreat into a laager of more tightly managed communication and reduced accountability. In Britain the opposite direction has been chosen: by being more open the police discover the value of an increased ability to enlist the public more closely into their decision-making and day to day activity. For instance in Derbyshire the response to a proposed merger with neighbouring police forces took the form of intense communication with the public through seminars, surveys, online and postal, and relevant website content. The resulting abandonment of the merger plans could be seen as being equally owned by the police force and the public. Also in Derbyshire, the response on the part of the police to a slightly higher than average fear of crime and perception that crime rates were high, has been to communicate better with the public to reassure them that, in fact, Derbyshire is measurably safer than otherwise comparable counties. The situation seems to be that police forces having perhaps somewhat unwittingly moved into an arena of greater openness for a complex of extrinsic reasons, now find themselves functioning highly effectively in the fresh environment and offering greater disclosure as a matter of operational choice.

Both Derbyshire and Leicestershire have comparatively recently created communications units dedicated to what is sometimes described as 'reputation management', but which is much more than this. These teams, of about a dozen in each case, include members with backgrounds in journalism, website management, public relations and marketing. In Derbyshire their work is seen as directed at three audiences: the general public; partner agencies such as the Criminal Justice system, ambulance and fire services, and local authorities; and interested parties, particularly the victims of crime. Their role includes creating a flow of positive stories for local press and media, to the extent that colleagues say that the Leicestershire communications staff 'keep the *Leicester Mercury* in business'. They also play a major part in facilitating communication face to face with the public, and drive aspects of more formal communication through the creation of publications and the incorporation of information technology applications.

FACE TO FACE COMMUNICATION

The whole pattern of police communication in both case study police forces is built on intense grassroots activity, which needs to be discussed before identifying the more formal face to face activities. Members of the communications teams stress the primacy of face to face communication between officers and the public in both disseminating messages and in obtaining insights and intelligence. This pattern of communication involves the use of meetings and publications, supported by the use of telecommunications media. In Leicestershire this is characterised as including police surgeries, visits to schools and youth groups, partner/agency work with bodies such as parish councils and Neighbourhood Watch groups, road shows as part of recruitment drives and many other informal and semi-formal means of contact. 'Street briefings' by officers on patrol are also a means by which exchanges take place with the public on matters such as priorities for surveillance and intervention. The Leicestershire force takes a stall in Loughborough market on a regular basis, so that the public can be informed about matters including crime prevention and personal safety. Identification and action on 'signal crimes' such as graffiti and littering, which contribute to public fear of crime, can be intensified through this type of contact.

Central to this approach is a policy of neighbourhood policing. Derbyshire, for instance, is divided into 103 neighbourhoods, each with its Safer Neighbourhood Team. These are the chief channel of communication to the public, and source of intelligence from the public. The Teams are made up of locally based police officers, Police Community Support Officers (PCSOs), special constables, park wardens and others. These teams, particularly their PCSO members, are seen as providing one of the few constant local presences (given the closure of village and neighbourhood shops and post offices and even the loss of many unprofitable local public houses). In Derbyshire the communications unit has four Community Officers supporting the work of the teams with materials such as posters, and publicity in the media.

A pattern of events, campaigns and other activities can then be constructed on this basis of widely distributed communication. For instance, Derbyshire launched its Operation Relentless in May and June 2007 as a high profile campaign to reduce the incidence of crime, improve detection rates and raise the visibility of policing activity. This involved a number of specific policing operations, for which an 8% reduction in crime, a 6% increase in detection and a 14% increase in arrests were claimed over a 61 day period. Associated with the campaign were 'Relentlessly Attacking Crime' seminars for the public, with demonstrations of police technology, dogs and the use of the police helicopter. In October 2007 the campaign was renewed with a focus on violent crime and anti-social behaviour; again with associated seminars in the north and south of the county. This combination of a targeted policing campaign linked to a publicity campaign seems to typify a major approach by police forces to integrating the role of communication in their activities. Other examples of localised communication are events such as Derbyshire's first Safer Neighbourhood Week, in which the public were invited to meet senior officers, visit police facilities and learn about the partnerships with other agencies. The police contributed to Liberation Day events, in which older citizens were encouraged to liberate themselves from the fear of crime.

More formal events share the same ethos of exchanging information and opinion rather than merely publicising the police position. For instance the budgetary consultation meetings held in Derbyshire in late 2006 involved the Chief Constable and financial officers making presentations about the budgetary options available in the spending of the next year's financial allocation. The cost of particular forms of operation was revealed and the ability to prioritise certain forms of policing in this financial context was set out. The large and attentive public audience was then able to ask questions make their points on expenditure priorities. In the same spirit, 25 'Have Your Say' events were held in 2006-7 at a range of venues including supermarkets, market places, university premises, hospitals, parks, libraries. Questionnaire responses on policing priorities and other matters were collected, and an adapted version of the budget consultation materials was used to extend the financial consultation process. At the same time,

> 'The events were designed to inform the public of Derbyshire about its police force. A wide range of literature relating to the Force and its performance was made available; annual reports, budget and Safer Neighbourhood team information was provided, as were small cards which were used to emphasise the force non-emergency number.' (Derbyshire Constabulary, 2007, p.9)

The integration of printed materials with the campaign of face to face events and meetings leads us, of course, to the specifically grey literature aspects of police communication.

GREY LITERATURE

The police both create and contribute material to grey literature. In the quite recent past the main print on paper item distributed by the Leicestershire force was the Annual Report. It was made publicly available in libraries, but today, of course, it is available through the force's website. Now there are glossy A4 publications led by *Your Police*. This is a 4 or 8 page newsletter, which carries a variety of short items illustrated with colour photographs dealing with changes and developments, policing successes, campaigns, and stories about police and community personalities. The same format and similar content is given a more local slant in *Local Policing in Rutland* (or any of the other areas served by the Leicestershire police). Finally there are neighbourhood newsletters with content aimed at specific districts, small towns and groups of villages, and presented as messages from the named neighbourhood beat team.

The pattern and content in Derbyshire is similar. *Policing Derbyshire* is described as the annual report. The 2005-6 edition is in newspaper format on glossy paper with colour illustrations. It adopts the friendly approachable style now obviously the norm for this type of material, breaking the statistics and statements down into separate news stories and also providing news from the various areas of the county. A noticeable item, headed 'Open and Transparent' draws attention to a step by step guide to the complaints procedures now available via the website. The same item gives basic figures for the complaints received during the year. There is also a smaller item to the effect that 390 Freedom of Information requests had been received. Derbyshire's *Safer Neighbourhood News* newsletters link to the activities arising from the priorities identified by the Safer Neighbourhood teams and also provide contact details and information on how to make complaints. The police force also feeds material into *Safer Derbyshire*, issued by the County Council. This glossy A4 publication, distributed free to every household, deals with a range of community and individual safety issues that include and go beyond policing.

If there is a consistent impression to be drawn from an examination of the grey publications of Derbyshire and Leicestershire, it is that friendly officers and police partners are out there, working in the community and available to give and receive help and information. In all of these publications the text and pictures are very professionally done and support each other in projecting a warm and positive message of openness.

TRANSPARENCY USING ELECTRONIC COMMUNICATIONS MEDIA

In the last three or four years the websites of police forces have come to be the focal aspect of their communication. They have their own 'Police' domain; their quality is excellent; and they carry high volumes of important information. They both substitute for a role that was entirely played by grey literature in the past, and at the same time generate and distribute a high volume of grey literature. They carry routine material such as the text of reports and newsletters, minutes of meetings; but also messages from beat officers and direct appeals to the public (which sometimes bypass the media and have achieved a certain success in dealing with cases). They are extensively used for recruitment activity; they are central to FOI response; and they project a consistently positive message about the force and its relationship with the public. For instance, the most endearing aspect of the Derbyshire website (www.derbyshire.police.uk) is its 'Young Space' area. This uses plain but lively language and high quality graphics (a plump crocodile police officer features) to tell children about police work, advise them on personal safety and amuse them with 'Fun Stuff'. The section that allows you to create a comic photo fit online is to be recommended for adults too. However, Devon and Cornwall Constabulary have gone a step further. A news release tells of the sticker albums issued in Brixham that enable children to approach officers and PCSOs to ask them for a sticker with their picture on it (Devon and Cornwall Constabulary, 2007).

One thing that all police websites do is to provide a starting point for FOI approaches by the public. The Derbyshire site, for instance has an FOI section that explains procedures and leads to the publication scheme through a 'List of Classes'. This is good but not one hundred percent obvious at first. Leicestershire (www.leics.police.uk) leads in through a section headed 'Library' which also effectively links into the publication scheme. This, in turn, offers one alternative that is 'A selection of booklets, leaflets and other material produced by Leicestershire Constabulary and its partner agencies, from crime prevention leaflets to guidance and advice'. To illustrate the value of this professionally produced, downloadable material, there is the powerful *Domestic Violence* publication (Leicestershire Constabulary, 2007). This deals in a very direct way with an issue too often neglected by society as a whole, let alone police forces. It provides police and other relevant local contact details, and is not only downloadable in pdf form, but large print, Braille and audio versions are also available.

There are naturally other implications of police electronic communication. For instance, police websites can provide links into various databases that are in the process of changing significant aspects of operational policing. Derbyshire maintains a database of people to whom they can target specific messages (Neighbourhood Watch groups, partner agencies, members of the police force, and particular targets of crime such as owners of Four Wheel Drive vehicles), through the 'Ringmaster' community messaging scheme. Some police forces provide a facility whereby victims of crime can key in a unique record number and check progress in the investigation of the case. Web access is widespread in Britain (although not at the levels of Scandinavia, North America and some other parts of the world) but there remains a need for easy public access. Derbyshire have experimented with kiosks for this purpose and have positioned one in the new Westfield Shopping Centre in Derby that opened on October 9th 2007. This will publicise Safer Neighbourhood activity and provide police contact details. It is just another example of the way that electronic information resources now sit at the heart of police communications activity.

THE ROLE OF FOI

Both Derbyshire and Leicestershire have Freedom of Information teams serving to manage records for retrieval and to handle responses to FOI enquiries. Under the FOI regime, it is required that a publication scheme is made available, revealing just what every 'public authority' already 'publishes' in some form or other. It is, in effect, the product of an organisation-wide audit of grey literature. Naturally the nation's police forces generate bodies of material that have a great deal in common. This made it possible for the Association of Chief Police Officers (ACPO) to coordinate the creation of publication schemes across the country. These provide a baseline document for responding to FOI requests, but they also contain lessons for the management of the organisation's total information resource. The question 'Can the record that provides the answer to question X be quickly identified and retrieved?' follows exactly the same line of approach that the question 'Can the published information that provides the answer to question Y be quickly identified and retrieved?' Whilst someone who has never worked with an organisation's grey literature might expect that the second question would be quite easy to answer, someone else who is experienced with this type of material will not necessarily be so confident.

Whilst this generalisation about grey literature may hold good, a similarly pessimistic view of the internal records of organisations (thoroughly justified by the neglect that is so common across many types of organisation) should not, and would not, hold good in relation to police forces. Police forces could not pursue effective investigations and bring cases to court successfully without excellent record-keeping. High profile failures based on the loss or inadequacies of records are subject to such media and public attention that the record-keeping activity is constantly reviewed and strengthened. Arguably it is more the human

failures of individual officers or units to follow established routines with regard to matters, including record keeping, that are responsible for most failures, not the systems themselves. Police forces as already information-intense organisations depend on good record keeping but FOI, by opening aspects of this to even more intense public scrutiny, has served to sharpen the quality of what they do. The number of FOI requests received by the two case study forces does not at present seem enormous (390 in Derbyshire during 2006), but other higher profile forces do report much greater levels. What is certainly the case is that both Derbyshire and Leicestershire note that FOI has resulted in increased workloads. Much of this will be for the preparatory work rather than the actual responses, but the total cost in salaries for a dedicated FOI team alone is substantial.

FOI is used by the general public, but the chief type of use is by the press and other media (an estimate for Leicestershire would be that something like 40-45% of all requests come from the media). The media have become very astute in their use of the Act and in avoiding the exemptions for which the law provides. Some of the requests are effectively 'fishing' for something that might make a story, and what is revealed may not mean as much as it at first seems. Take, for instance a *Derby Evening Telegraph* story headlined 'Dozens of young people given a licence to shoot' (Oakes, 2007). This revealed that of 436 firearms licences issued in the previous year in Derbyshire, 33 were for people who were under 17. The headline and even more the subheading 'We force police to reveal number of shotgun permits for under-17s'give the impression that this might be an attempted police cover-up of a dangerous gun problem, the story itself shows that the permits were for gun use under supervision by farmers' children and potential sporting shooters. FOI requests are answered literally from the records available and the answers may not actually address the question in the enquirer's mind. For instance an enquiry about seizures of explosives might produce a high and rather frightening figure that when broken down consists chiefly of fireworks and gun cartridges – not the kind of explosives that might indicate potential terrorist activity. An enquiry as to how many times there was a police officer in a particular village might only include formal, logged visits and fail to reveal many other visits either recorded incidentally in other files, or not recorded at all.

However, FOI is a significant creator of transparency. Some forces already publish the responses they make to every FOI request, via their websites, thus moving the information into a grey status. This clearly sets a standard that might seem likely to become accepted practice nationally. Whilst FOI activity continues and increases, it is important to note that any police force deals with countless thousands of enquiries as a matter of routine ('If you want to know the time ask a policeman' as the old music hall song says). It is probable that in the current climate of openness the number of significant enquiries that are now answered as routine (when they might have been avoided in the past) and never reach the status of FOI request is considerable. Leicestershire receives about 48,000 media calls per year, to which can be added the FOI requests, but which fails to give any measure of the volume of everyday queries to police personnel. Perhaps the only certainty is that public interest in police work will not diminish. The press will continue to contact the police for stories, and the experience of countries such as Canada with longer-established FOI regimes suggests that the numbers of FOI requests will continue on an upward trend.

CONCLUSION

Police communication is important: people need open, responsive and accountable police forces. In the UK a combination of factors, including Freedom of Information legislation, have in a remarkably short time created a new and well rooted ethos of openness. There is now an acceptance that high levels of disclosure are, and should be, the norm. This has brought increasing volumes of information that were previously unlikely to be disclosed into grey status. Not only is information prepared by professional communicators for publication and broadcasting, but police websites also communicate and act as resource banks for grey literature. At the same time the volume and quality of traditional grey publications has increased. Poor production quality and unenthusiastic distribution might have common in the past, but today the effort and inspiration displayed suggests genuine intent to communicate. It might be possible to argue that this flow of slick, professional communication is another example of twenty first century media superficiality. Looked at in a context that includes FOI and the general enhancement of police information handling that is in process, this accusation would not hold much credibility.

The public can undoubtedly benefit from what they are able to learn about police activity, and the relations between people and their police forces can also only benefit. There are however costs. Professional communication costs money and an FOI regime costs money. Police budgets are not infinitely expandable and if openness were to reach costs levels that threatened operational policing it would not only be police officers who were worried. There is a ratio that balances the ethos of openness with the resource demands that it makes. At present all concerned seem to accept that the ratio is in an acceptable balance. The future of policing in the UK requires that this continues to be the case, even if the costs continue to rise. The will of governments and the support of voters are needed for openness in policing to be maintained and, when necessary, expanded.

NOTES

1. Derbyshire case study material acquired from interview with Jonathan Leach, Head of Corporate Communications (13th Sept 2007) and as participant observer at Budgetary Consultative meeting (1st Feb 2007) and 'Operation Relentless' Seminar (18th Oct 2007).
2. Leicestershire case study material acquired from group interview with Emma Rogers, Head of News; Anne Chafer, Information Manger; and Linda Dempsey, Information Assistant, Data Protection & Freedom of Information (26th Sept 2007).

REFERENCES

Bichard, M. (2004) *Bichard Inquiry.* London: Home Office. Available at: http://www.bichardenquiry.org.uk/report/ [Accessed 9 October 2007].

Brodeur, J.-P. and Dupont, B. (2006) Knowledge workers or "knowledge" workers? *Policing and Society*, 16 (1), 7-26.

Comrie, M. and Kings, E. (1975) *Study of Urban Workloads: Final Report.* London: Home Office Police Research Services Unit.

Derbyshire Constabulary (2007) *Have Your Say 2007: Report*. Butterley, Derbys: Derbyshire Constabulary and Derbyshire Police Authority.

Devon and Cornwall Constabulary (2007) Building Safer Communities Together. Available at http://www.devon-cornwall.police.uk [Accessed 11 October 2007].

Ericson, R.V. (1982) *Reproducing Order: a Study of Police Patrol Work.* Toronto: University of Toronto Press.

Ericson, R.V. (1993) *Making Crime: a Study of Detective Work.* 2nd ed. Toronto: University of Toronto Press.

Ericson, R.V. and Haggerty, K. (1997) *Policing the Risk Society.* Toronto: University of Toronto Press.

Home Office (2004) *Building Communities, Beating Crime: A Better Police Service for the 21st Century.* Cmnd. 6360. Available at: http://www.homeoffice.gov.uk/documents/wp04_complete.pdf [Accessed 9 October 2007].

Leicestershire Constabulary (2007) *Domestic Violence*. Leicester: Leicestershire Constabulary. Available at www.leics.police.uk [Accessed 10 October 2007].

Luen, T.W. and Al-Hawamdeh, S. (2001) Knowledge management in the public sector: principles and practices in police work. *Journal of Information Science*, 27 (5), 311-318.

National Centre for Policing Excellence (2005) *Code of Practice on the Management of Police Information.* London: Home Office.

National Centre for Policing Excellence (2006) *Guidance on the Management of Police Information.* London: Home Office.

Oakes, C. (2007) 'Dozens of young people given a licence to shoot'. *Derby Evening Telegraph*, 27th Sept.

Parliament (1999) *Modernising Government.* Cmnd. 4310. London: The Stationery Office. Available at: http://www.archive.official-documents.co.uk/document/cm43/4310/4310.htm [Accessed 9 October 2007].

Royal Commission on Criminal Justice (1993) *Report.* Cmnd. 2263. London: HMSO.

Syson, H. (2007) *Knowledge Sharing within the UK Police Service.* Loughborough University. Unpublished MSc dissertation.

A Scholarly communication perspective of Grey literature in Library and Information Science education

Debbie L. Rabina
Pratt Institute, School of Information and Library Science (United States)

Abstract
The purpose of this article is to inform library and information science (LIS) educators of the status of grey literature among LIS students in the United States. More specifically the article examines what students enrolled in ALA-accredited library and information science programs know about grey literature and where they learned about it. Masters-level students at three North American LIS programs answered a four-question survey. Findings revealed that knowledge about grey literature is gained across the curriculum with general reference sources being the course most likely to address grey literature. Findings also indicate that knowledge about grey literature is more intuitive and anecdotal than systematically acquired. In addition to the implications of the results, the relationship among grey literature, the library profession, and library education is discussed. The article concludes with suggestions for further analysis and research.

Literature review and justification
The most-cited definition for grey literature is "that which is produced by government, academics, business, and industries, both in print and electronic formats, but which is not controlled by commercial publishing interests and where publishing is not the primary activity of the organization" (Farace, 1997). ODLIS (Online Dictionary of Library and Information Science) provides a slightly different definition of the term: "Documentary material in print and electronic formats, such as reports, preprints, internal documents (memoranda, newsletters, market surveys, etc.), theses and dissertations, conference proceedings, technical specifications and standards, trade literature, etc., not readily available through regular market channels because it was never commercially published/listed or was not widely distributed." While definitions proliferate, there is agreement on the main characteristics of grey literature: they are materials that are published not for profit and, as a result, typically not marketed or distributed by commercial publishing organizations (Mackenzie Owen, 1997). In other words, grey literature is discussed in terms of its origins, its methods of dissemination, or both.

The research conducted by Sulouff et al. (2005), whose paper is most closely related to the theme of this study, points out that grey literature "takes different forms in different departmental settings" so that a working definition is often based on circumstance. The library sector carries responsibility for the management and processing of grey literature. This role is acknowledged by several researchers (Mackenzie Owen, 1997; Sulouff et al., 2005) although they have written largely about the role that librarians take with regard to grey literature, but little about how librarians learn about grey literature. The role of librarians is described as promoting dissemination and use of grey literature through cataloging, searching, archiving and preservation (Mackenzie Owen, 1997). Gelfand believes that these roles, at least with regard to grey literature, are learned on the job: "training and bibliographic familiarity… does not follow a curriculum or a set of readers of textbooks, but instead studied by doing" (Gelfand, 1998).

Research regarding grey literature in library and information studies has focused on case studies, such as that by Aina (2000), more than on grey literature in library and information science education. An environmental scan, described in more detail below (note 10), supports Gelfand's view that training in grey literature is mostly field, and not curriculum, driven. With regard to formal instruction of grey literature in library and information science programs in the United States, the only full scale framework identified is Dominic Farace's distance education course offered at the University of New Orleans, though one must note that the course is offered as part of a school library media program in the department of education, and not in a Master's of Library Science program[1].

Library and information professionals are a vital link in the chain that makes grey literature available to researchers, students and the interested public. While on-the-job training is invaluable, certainly the purpose of graduate-level training is that professionals are hired with some baseline knowledge that they bring to the work place upon graduation. Courses that educate future librarians in areas relating to grey literature are critical training ground if awareness to grey literature is to increase.

The purpose of this paper is to identify what students currently enrolled in MLS programs know about grey literature and where they learn it. Once we have a clearer picture of the training currently available, we can open a discussion between library professionals, library educators and library students to determine how library education can best assist in meeting the needs of the current workplace and use LIS education to strengthen the relevance of current graduates to the workplace.

Methodology and data collection

Since the purpose of this study is exploratory, a survey of masters-level library and information science schools students in the form of an online questionnaire was deemed the most appropriate form of data collection. The survey was conducted in October 2007. The three master's programs selected—Wayne State University in Michigan, Rutgers University in New Jersey and Pratt Institute in New York City—were determined by the researcher to represent a range of LIS programs in the United States. While all three are library and information science schools, there are notable differences between them in character and emphasis.

Rutgers University, School of Library, Communication and Information Studies, is part of the iSchools caucus[2] and their research and educational goal, as described on their website, is to study the digital revolution and related areas on human organizations, social organizations and institutions[3]. While Rutgers-SCILS educates at all levels, from undergraduate to Ph.D. and in several disciplines (in additional to library and information science, degrees are also offered in journalism and communication), only students studying for a master's degree in library and information science were selected for this study.

Wayne State University, in Detroit, Michigan, is what may be considered a traditional library and information science program. The program does not have a Ph.D. program and serves MLS students and certificate students, with a great emphasis on skills that will make students attractive to the job market[4].

Pratt Institute, in New York City, part of an institution of higher education whose focus is primarily on art and architecture, attracts many masters-level students – there is no Ph.D. program—who are interested in aspects of cultural informatics as a cross section of information, design, and globalization[5].

Three separate online surveys and one hand-distributed survey were conducted using a commercial survey tool. The surveys were distributed through the listserv of each department and addressed to the students of the individual departments. The survey contained 4 brief questions (see appendix I) and took no longer than 5 minutes to complete.

While the survey questions were identical on all surveys, the question relating to courses in each department were tailored to the course names and corresponding numbers in each department. Results were then aggregated into a single file to allow comparison.

The survey ran for 10 days in early October 2007. In all, 294 students responded to the survey: 101 from Wayne State (17% of Wayne State LISP 570 students), 82 from Rutgers (24% out of Rutgers SCILS MLS 338 students) and 111 from Pratt (31% of Pratt-SILS 350 students). Data was compiled and analyzed using SPSS, a statistical analysis software package.

Limitations of surveys as a data collection method are inherent in the instrument, mostly that results are self reported and could be skewed by intentional deception, misinterpretation of the questions, and a desire to please the researcher[6]. To avoid these limitations to the greatest degree possible, the questionnaire was tested for reliability in a pilot study conducted with a small group of students during July 2007, and the questions revised based on the group's feedback. In addition, of the 111 surveys completed by Pratt students, 37 were distributed in class rather than online. The two groups were compared to see if statistically significant differences existed between the groups, and none were found.

Findings

The survey was answered by 294 respondents from three Masters' of information and library science departments in the US; with Pratt students being the largest group represented, Wayne State second and Rutgers third.

Table 1: Number of respondents and their institutional affiliation

	Number of respondents	Percent of respondents[7]
Wayne State	101	34%
Rutgers	88	29%
Pratt	111	37%
Total	294	100

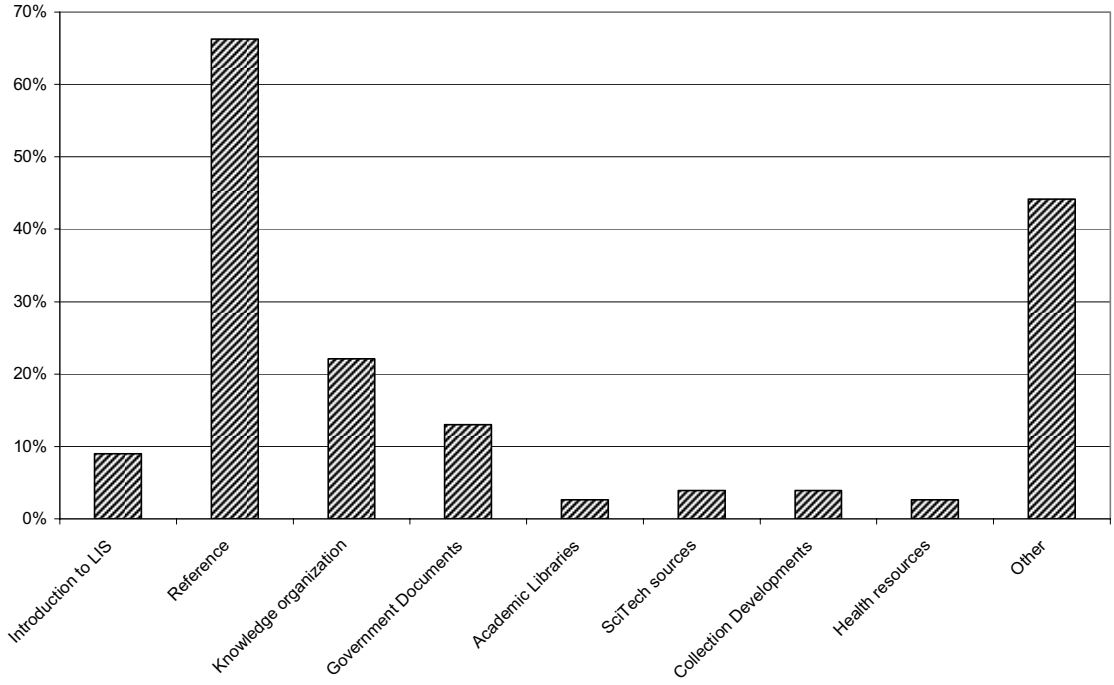
Results revealed that the percentage of respondents who heard of grey literature, was just over a quarter of all respondents, 26.2%.

Table 2: Number of students who heard of grey literature

	Number of all	Percent of all	Percent within school
Wayne State	15	5.1%	17% of students in program
Rutgers	21	7.14%	24% of students in program
Pratt	41	13.9%	31% of students in program
Total	77	26.2%	

Of the 77 students who indicated that they had heard of grey literature, 63 indicated in which course or courses they learned of the term. Several students selected more than one course. Most respondents heard of grey literature in reference classes, with knowledge organization and government documents being the next two categories. It is important to keep in mind that the question that generated these results was tailored to the course offerings in each school, and that not all schools offer the same selection of courses, nor are the courses given the same titles. For example, Rutgers-SCILS does not offer a course in Introduction to LIS, while both Wayne State and Pratt do.

Chart 1: Courses that informed students of GL



Students were given nine statements (see appendix) and asked to rate each one in terms of how well the statement describes the term Grey Literature. Responses were on a Likert scale with 5 meaning the statement describes Grey Literature very well and 1 meaning the statement does not describe it well.

Table 3: suitability of statements to describing Grey Literature

Very well	Statement	Not well
19	*1. Grey literature* describes materials of unknown origin (where the author or publisher can't be identified)	32
2	*2. Grey literature* is similar to open access journals	40
5	*3. Grey literature* refers to materials guarded by institutional gatekeepers who deny access to them	49
1	*4. Grey literature* refers to materials stored in dark archives that are intended for long term preservation	79
39	*4. Grey literature* are materials not indexed by commercial indexers	14
20	*5. Grey literature* is government information that is not available in the *Catalog of Government Publications*	30
31	*7. Grey literature* describes materials published by non-commercial publishers	18
24	*8. Grey literature* describes materials not available in OPACs	18
25	*9. Grey literature* describes materials not picked by my commercial search engines (such as Google and Yahoo)	25

Discussion

The results suggest that the case of grey literature in library and information science departments is haphazard and that there is no pattern to what, where, or how students know of grey literature. Nonetheless, a closer examination reveals some interesting findings.

Duration in MLS programs made no difference in perceptions of the nature of grey literature, with the exception of statement 4 ("grey literature refers to materials stored in dark archives that are intended for long term preservation") which incoming students (Fall 2007) were significantly *more* likely than others to indicate as an inapt description. It thus appears that duration in the MLS program is not an indicator of having heard of grey literature or of being able to describe it accurately.

Of the 77 students who indicated hearing the term, 74 ranked the statements provided in the questionnaire, but in addition to them, 66 students who claimed not to have heard the term grey literature, also ranked the statements. A statistical comparison between the two groups found no significant difference in the way each group perceived grey literature. This finding alone may indicate that the scope and depth of knowledge acquired throughout the MLS experience, allows students to make informed judgments regarding the accuracy of the statements provided in the questionnaire.

Statistically significant results regarding the nature of grey literature were found in the rating of five of the nine descriptions. Students perceive in descending order that grey literature are materials not indexed by commercial indexers, that grey literature describes materials published by non-commercial publishers and that grey literature describes materials not available in online public access catalogs (OPACs). On the other end of the spectrum, students do not identify grey literature as materials stored on dark archives, nor as materials whose author or publishers are unknown.

The question is, of course, did students get it right? Students' perception of grey literature as not indexed and published by non-commercial publishers is within the scope Farace's narrow definition of grey literature. With regard to the third significant finding, that grey literature describes materials not found in OPACs, there is no definition that supports this perception. On the other end of the spectrum, students feel that the statements that least accurately describe grey literature are those that refer to it as materials stored in dark archives and materials whose author or publisher are unknown. While definitions do not for the most part describe something in terms of what it is not, the two statements students identified as describing grey literature least well are closely linked to other forms of materials that have weak bibliographic control, namely, government documents and orphan works.

When attempting to learn the courses in which students learned about grey literature, general reference courses received the highest scores both overall and within each school. While this finding lacks ambiguity, it is worthwhile to note that previously gathered evidence seems to indicate that the library professional most likely to deal with grey literature in libraries are serials librarians, collection developers or technical services librarians, and only then reference librarians[8].

Conclusion and recommendations

Grey literature and emerging publishing models are changing the role of librarians. Traditional library roles which were for the most part departmentalized, are converging, and grey literature is a good example of the new playing field of libraries.
While grey literature has a following among academic librarians, this study finds that Gelfand's assertion that it has gained a new visibility may be a bit optimistic (Gelfand, 1998), as it seems that today, nearly a decade later, only a quarter of students currently enrolled on library and information science program even heard of the term.
Findings indicate that grey literature is covered - or more accurately, mentioned - in many different courses in MLS programs[9]. This treatment of grey literature across the curriculum is appropriate to demonstrate the various aspects, treatments and handling of grey literature in libraries. Curriculum developers would do wisely to continue and expand this approach.
To get a fuller picture of the status of grey literature in LIS education, a more comprehensive examination of course syllabi should be conducted[10].

Appendix

Survey on Grey Literature

Dear student,

Please take a few minutes to answer this short survey on Grey Literature. Results will be used as part of a study on library education, presented in a scholarly conference, and made available to students. No personal data is collected and respondents remain anonymous.

If you have any questions, please feel free to contact me at drabina@pratt.edu.

Thank you for your cooperation.

D. Rabina

1. In the course of your studies, have you come across the term Grey Literature?
   - ☐ Yes
   - ☐ No

2. To the best of your recollection, in which course or courses have you heard the term (check all that apply) [*note: list of courses was tailored to each school's offerings*]
   - ☐ Knowledge organization
   - ☐ Reference
   - ☐ Other
     _____
     _____
     _____

3. When did you begin your MLS studies?
   - ☐ Fall 2007
   - ☐ Spring 2007
   - ☐ Fall 2006
   - ☐ Spring 2006
   - ☐ Fall 2005
   - ☐ Spring 2005
   - ☐ Other _____

4. **Which of the following statements best describes the term Grey Literature?**

	Very well 5	4	3	2	Not well 1
Grey literature describes materials of unknown origin (where the author or publisher can't be identified)					
Grey literature is similar to open access journals					
Grey literature refers to materials guarded by institutional gatekeepers who deny access to them					
Grey literature refers to materials stored in dark archives that are intended for long term preservation					
Grey literature are materials not indexed by commercial indexers					
Grey literature is government information that is not available in the *Catalog of Government Publications*					
Grey literature describes materials published by non-commercial publishers					
*Grey literatur*e describes materials not available in OPACs					
Grey literature describes materials not picked by my commercial search engines (such as Google and Yahoo)					

Bibliography

Aina, L.O. (2000), Grey literature and library and information studies: A global perspective. *The International Journal on Grey Literature 1* (4), 179-182.

Farace, D. Third Annual Conference on Grey Literature, Luxembourg, November 1997.

Gelfand, J. (1998). Teaching and exposing grey literature: What the information profession needs to know – examples from sciences. *Collection Building 17* (4), 159-166.

Mackenzie Owen, J. (1997) The expanding horizon of Grey Literature. In *Proceedings GL3 : Third International Conference on Grey Literature*, Luxembourg (Luxembourg).

Sulouff, P., Bell, S., Briden, J., Frontz, S., Marshall, A. (2005). Learning about grey literature by interviewing subject librarians: A study at the University of Rochester. *College and Research Libraries News 66* (7) (access online Nov. 11, 2007 http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2005/julyaugust05/learngreylit.cfm)

References

1 Course syllabus: http://www.greynet.org/images/Syllabus_Grey_Literature.pdf (access Nov. 24, 2007)

2 iSchools http://www.ischools.org/oc/ (accessed Nov. 24, 2007)

3 About SCILS http://www.scils.rutgers.edu/about-scils/about-scils.html (accessed Nov. 24, 2007)

4 Wayne State – LISP http://www.lisp.wayne.edu/ (accessed Nov. 24, 2007)

5 Pratt Institute – SILS http://pratt.edu/~infosils/sils.html (accessed Nov. 24, 2007)

6 Thus, the high participation rate of Pratt students may be attributed to the fact that the researcher is a faculty member of that department.

7 All percentage points in this paper are rounded to the closest hundredth of a percent.

8 For more details see Association of Research Libraries, Scholarly Communication Education Initiatives, August 2007. http://www.arl.org/bm~doc/spec299web.pdf (access Nov. 24, 2007). E-mail posting from Serials Listserv (SERIALIST) seems to support the view that serials librarians and collection developers have responsibilities for non traditional sources such as grey literature.

9 An examination of LIS syllabi was attempted in order to identify courses that covered grey literature. Because many universities do not make their syllabi available on the free Web, a comprehensive examination was not possible. However, the syllabi that were examined represent a wide range of LIS course offerings. As mentioned in the Literature Review, only one course was found that is dedicated to grey literature (New Orleans http://www.greynet.org/images/Syllabus_Grey_Literature.pdf access Nov. 24, 2007). Other include general reference (Louisiana (http://slis.lsu.edu/syllabi/7002.pdf accessed Nov. 24, 2007); records management (Chapel Hill http://www.ils.unc.edu/courses/2007_spring/inls525_001/inls525-spring2007-syllabus.htm accessed Nov. 24, 2007), health information sources (Texas A&M edhttp://www.library.ucsb.edu/istl/00-fall/article4.html accessed Nov. 24, 2007), SciTech information sources (Clarion http://www.clarion.edu/libsci/ls556_041_w1.htm and University of Texas http://www.gslis.utexas.edu/~l382l4ne/syllabus.html both accessed Nov. 24, 2007), and collection development (Kentucky http://www.uky.edu/CIS/SLIS/syllabi/659_dc.pdf, Simmons http://www.google.com/search?q=course+grey+literature+syllabus+lis+site:edu&hl=en&lr=&as_qdr=all&pwst=1&start=10&sa=N and Hawaii http://www2.hawaii.edu/~wertheim/615S07Syllabus.pdf all accessed Nov. 24, 2007)

10 See note 10 supra

Grey Literature: A Pilot Course constructed and implemented via Distance Education

Dominic J. Farace and **Jerry Frantzen**
GreyNet, Netherlands

Joachim Schöpfel and **Christiane Stock**
INIST-CNRS, France

Abstract
Over the past years, training courses, guest lectures, seminars and workshops have been organized by information professionals on the topic of grey literature. Most of these endeavours have undoubtedly had some impact on this field of information, however difficult it would be to measure. Many of these course lectures were given within existing programs and may have only been one-time offerings. At the Eighth International Conference on Grey Literature (New Orleans, 2006), two of the authors in this pilot study participated in a roundtable on curriculum development and grey literature. For their part, it was more important to find a proper structure within which to further construct an accredited course than to simply provide an inventory of *ad hoc* training courses or workshops, which deal/dealt with grey literature. The intended structure would have to incorporate the expertise of a number of stakeholders in order to guarantee potential students course credit, access to courseware and resources, qualified instruction, etc. Early on, it became evident that these stakeholders need not be physically present within one particular academic institution, but could rather be brought together in a joint venture by way of distance education. This paper will focus on the stakeholders in the pilot program and the specialization of each, as well as the students who are profiled and the knowledge and skills from which they would benefit. Built into the pilot is the maintenance of an ongoing log that would capture the pilot courses' development and progress, facilitate a SWOT analysis, enable comparison with other distance education courses in the LIS (Library and Information Studies) sector, and ultimately substantiate this course offering beyond a pilot phase to academic institutions with degree programs in information and other related fields on undergraduate and graduate levels.

Background
At the Fourth International Conference on Grey Literature (Washington D.C., 1999), Julia Gelfand presented a paper dealing with grey literature and distance education[1]. At the Sixth International Conference on Grey Literature (New York, 2004), the results of an online survey in which 102 respondents participated not only indicated that an increasing number information professionals involved in grey literature were also involved in teaching and instruction but also they were in agreement with the statement (one of the 22 items on the questionnaire) that grey literature warrants a special field in information studies[2]. At the Eighth International Conference on Grey Literature (New Orleans, 2006), during a Roundtable on Curriculum Development[3] the content for a proposed course on grey literature and its subsequent headings was drafted. And now, this present study seeks to harness the results of the above mentioned conferences incorporating them in a proposal for a pilot course on grey literature that would be offered for full academic credit.

Premise and Goal
If Grey Literature constitutes a field in Library and Information Studies, then it must be offered to LIS students as an accredited college course. Given the state of the art in grey literature, Distance Education now offers a viable means of implementing an accredited course on Grey Literature

Method of Approach
In order to proceed further based on the premise and goal set forth in this paper, it was necessary to establish the status quo found in teaching grey literature to students, to confirm the authors' position that there was no course for credit entitled 'grey literature' or where 'grey literature' was the primary topic, and finally to construct and implement a pilot course on grey literature that would be offered at an accredited academic institution via distance education. The steps or phases in our approach were not in strict sequential order but were rather carried out by the team of authors/researchers simultaneously and monitored throughout the study. In order to establish what we refer to as the status quo, a survey was devised and carried out via an online, bilingual questionnaire. The eight item questionnaire in English and French was made available on GreyNet's website and was further distributed via channels belonging to INIST and GreyNet. The choice of these channels was based on the organizational affiliations of the authors. The duration of the online questionnaire was one month (September 10th through October 10th 2007) after which the responses were compiled and processed. The eight item questionnaire can be found as an *appendix* to this paper. The literature search carried out in this study was not intended to be exhaustive nor comprehensive. The authors were less interested in providing an inventory of courses that had been offered dealing with grey literature than confirming their premise that no course for credit

entitled grey literature was part of an existing academic program. Examples from the literature search linked to findings from our own questionnaire offer a control for our study. Finally, a pilot course for credit entitled 'grey literature' was constructed and submitted in the spring of 2007, and was later approved by the University of New Orleans Distance Education Department for the fall 2007 semester.

Results of the Survey

A total of 19 online questionnaires were completed and returned, 18 of which were valid. Based on the domain name in the email address, the respondents were from 10 countries in North America and Europe. Fourteen of the respondents were from universities, two from LIS colleges, and two from research institutes. Thirteen of the respondents held the status of an instructor, two were former students, and three did not respond to the item. Instruction was either given or received on a doctor's level for five of the respondents, thirteen on a master's degree level, eight on a bachelor's degree level, and one respondent left the item blank. Six of the 18 respondents had completed a thesis or research project on grey literature. Nine of the respondents participated in one or more of the conferences in the GL-Series, while the other nine had not. Examining further the results of the survey, it became clear that another item should have been added to the questionnaire dealing with the duration of the course and the percentage of which that dealt with grey literature?

Looking at the course titles (item 5), only one explicitly refers to the production of a document. However, when we examine the content of the courses (item 6), we find that others do mention aspects dealing with the production side of grey literature. Nevertheless, these were far outnumbered by aspects involved in the search, access, and distribution of grey literature. We also find that the content of the courses (item 6) were either very specific to one aspect of grey literature *e.g.* collection development or searching for grey literature; while other courses treated a number of aspects of grey literature - ranging from its history and definition to problems and challenges it faces.

It became apparent that the production and publishing side of grey literature was significantly less a part of the content taught than the processing, distribution, and access to grey literature. This apparently sidesteps the accepted definition of grey literature known as the Luxembourg convention (1997) "that which is produced on all levels of government, academics, business and industry in print and electronic formats not controlled by commercial publishers"[4]… "*i.e.* where publishing is not the primary activity of the producing body"[5]. This clarification was added after the Sixth International Conference on Grey Literature in New York (2004) and was based on findings from two research projects in which GreyNet participated. One was an online survey, where 77% of the respondents agreed that the Luxembourg definition on grey literature still holds with or without modification[6]. And, the other a citation analysis in which the authors "… find that the term 'commercial' in the Luxembourg definition of grey literature needs to be further elaborated in order to better serve its purpose"[7]. From the current authors' perspective, a course on grey literature should pay equal attention to the production side of grey literature: how is it produced, who are the producers and publishers, the review process, rights issues, marketing strategies, as well as economic models available for grey publishers. Instructors should make explicit to their students that the more grey literature adheres to common and shared standards in its production and publication (e.g. its appearance, value added metadata, and quality control), the easier it is to access and retrieve it. And, in so doing, the misnomer that grey literature is 'unpublished'[8] or 'non-published' would be countered and adequately contested.

We also find that instructors use intermittently less defined terms such as 'fugitive literature', 'ephemera', or 'non-conventional literature', when they are actually referring to grey literature. It is almost as if LIS instructors are unaware of research that has been carried out over the past 15 years by hundreds of researchers and authors in the field of grey literature. Results that have been published in a variety of document types: conference papers, proceedings, journal articles, reports, etc. in both print and electronic formats. This might even lead us to suggest that a course on grey literature may not only be worthwhile for LIS students but also would have a place for LIS instructors via continuing education and teacher accreditation programs. However, this would take us well beyond the scope of our present study – the construction and implementation of a pilot course on grey literature via distance education.

Pilot Course for Credit on Grey Literature

During the first quarter of 2007, GreyNet established contact with the Provost at the University of New Orleans (UNO) and EBSCO Publishing. Together, these three organizations would comprise the major stakeholders in the pilot course on grey literature. UNO would provide an established distance education program, registration for potential students, and subsequent accreditation. EBSCO would provide online access to the full-text journal articles comprising the required readings for the course, and GreyNet in cooperation with INIST, one of its Associate Members, would carry out the selection and development of resources and courseware required for the pilot and the instruction of students. In the spring of 2007, the pilot course entitled 'Grey Literature' was submitted to the UNO Director of Distance Education; and it was later approved as a course offering for the fall 2007 semester.

System Software, Courseware and Assignments

UNO provided access to Blackboard (Bb)[9], the system software that enabled email connectivity, a dedicated announcement board, document archive, up-to-date information on student enrolment, the

grade book, and other useful functions. GreyNet then in turn used Bb to post and archive the instructor's CV, the Syllabus including the course title and description, information about the course assignments and assessment, a weekly outline of the semester, and due dates for student assignments. The required reading list (e-Reader) was also posted. The e-Reader was divided into five sections, each section dealing with a particular aspect of grey literature. Each section contained four articles. The full-text of all 20 journal articles was available via EBSCO's LISTA-FT (Library, Information Science & Technology Abstracts) full-text database[10]. This resource was available to all UNO students both on and off campus. In addition, an Appendix to the e-Reader provided links to other resources in grey literature.
All in all, the course consisted of three assignments: an open-ended test based on the assigned readings, a case study, and the design of research proposal on grey literature. Information sheets pertaining to the case study and research proposal were also posted and archived in Bb. The results of each the three assignments weighed equally for a third of the semester grade. Each assignment covered a five week period in the semester; and, by the 16th week all assignments had to have been submitted to the instructor for assessment, grading, and final posting.

Student Results
During the initial period of enrolment, 14 students registered for the course on grey literature (EDLS 4990) in the fall 2007 semester. They came primarily from the departments of general education and library science. Nine of the students completed the course, while the remaining five dropped within the grace period. Six of the students who completed the course were senior undergraduates, one junior, and two sophomores. It was originally the understanding that only upper-level undergraduates *i.e.* juniors and seniors would be admitted to this course. Interesting to note that the two sophomores received the lowest grades averaging 72.5% compared with the overall grade average of 76.9%. On the assigned readings test, the students had a 76.4% grade average, on the Case Study an 81.3% grade average, and on the Design of a Research Proposal, a 73.1% grade average. The highest student grade for the semester was 95% and the lowest was 71.7%.

Course Results
The Syllabus and 16 week planner, which was posted to Bb in August, remained unchanged throughout the semester. Three credit hours were given to the nine students who completed the course. The withdrawal rate for this course was 35.7% compared with an average of 20% for distance education courses. All students who registered for this course were already previously enrolled at UNO. Three students outside UNO inquired about possible enrolment, but their registrations were never recorded. The course instructor had ca. 125 email communications with students during the course and ca. 80 email communications with UNO administrators. The instructor posted 14 messages to Bb during the course of the semester – nine of which were permanent postings and five of which were temporary ones. The instructor personally met once during the semester with the Director of the UNO Distance Education program in New Orleans and once with the co-authors in the pilot study. Both meetings were used to discuss the progress of the course and the approach that would be used to evaluate the course and its findings.

Some Closing Remarks
Our survey and literature search affirm that no prior course-for-credit had been either given or taken on grey literature. Based on the results that are available, the pilot course met with measurable success for all of the stakeholders. While the results from the student evaluation forms and the post-semester assessment by UNO faculty are still forthcoming, we hope they will provide answers to questions that arose during the course of the semester and which are still unanswered: Should the course be available to General Education as well as LIS students? Should a student who completes the course be allowed to carry out his/her research proposal for further academic credit? Can this course be marketed to other LIS Colleges and Schools? Is the content and structure of the course viable beyond a Distance Education Program?
In fine, based on the limited results of our pilot, Distance Education does provide an adequate structure for a college level course on Grey Literature. The results from this pilot will no doubt contribute to the development of course content as well as course instruction on grey literature. And, emphasis should be placed on securing investments that have already been made by the stakeholders in this pilot.

Appendix:

BILINGUAL SURVEY ON GREY LITERATURE IN EDUCATION

This survey seeks to evaluate the place grey literature holds today in courses offered in Library and Information Science. It is carried out in advance of the Panel Session on Education that will be held during the Ninth International Conference on Grey Literature this December in Antwerp, Belgium. The survey is carried out via the distribution channels of INIST and GreyNet and should take no more than 5 minutes to complete. Respondents will have until October 10, 2007 to complete the online questionnaire, below.

L'enquête suivante cherche à évaluer la place de la littérature grise dans l'enseignement des sciences de l'information. Elle est menée par l'INIST et Greynet dans le contexte d'une session sur l'éducation lors de la 9ème Conférence internationale sur la littérature grise à Anvers/Belgique en décembre 2007. 5 Minutes devraient suffire pour remplir le questionnaire.

1. Your email address?
   Votre adresse mél?

2. Are you a Student/former Student or an Instructor?
   Etes-vous un étudiant, un ancien étudiant ou un enseignant en IST?

3. On what level(s) of education are/were you engaged (e.g. BA, MA, PhD, etc.)?
   Quel est le niveau d'études suivis / enseignés (licence, mastère, thèse)?

4. What is the name of your academic institution?
   Quel est le nom de votre université?

5. What is the title of the course in which grey literature is covered?
   Quel est le titre du cours qui parle de la littérature grise?

6. Can you briefly describe the content dealt with on grey literature?
   Décrivez-nous brièvement le contenu de l'enseignement sur la littérature grise?

7. Did you complete a thesis or research project on a subject dealing with GL?
   Votre mémoire/thèse ou projet d'étude porte-t-il sur un sujet touchant/incluant la littérature grise ?

8. Other Comments?
   Autres commentaires?

Thank You / *Merci* 10/SEPT/07

References

[1] Gelfand, J. (2000), Grey Literature: Plausible Applications for Distance Education Practices. – In: GL'99 Conference Proceedings: Fourth International Conference on Grey Literature "New Frontiers in Grey Literature" 4-5 October 1999, Washington D.C. - Amsterdam : GreyNet, January 2000, pp. 1-11. – (GL-conference series, ISSN 1386-2316 ; No. 4). - ISBN 90-74854-24-9

[2] Boekhorst, A.K., D.J. Farace, J. Frantzen (2005), Grey Literature Survey 2004: A research project tracking developments in the field of grey literature. – In: The Grey Journal: An International Journal on Grey Literature, TGJ Volume 1, Number 1, 2005, pp. 41-50. – ISSN 1574-1796

[3] D.J. Farace and J. Frantzen, ed. (2006), GL8 Conference Program and Abstracts:
Eighth International Conference on Grey Literature : Harnessing the Power of Grey, 4-5 December 2006, New Orleans. – Amsterdam : TextRelease, December 2006, pp. 6-7. - (GL conference series, ISSN 1385-2308 ; No. 8) . - ISBN 90-74854-07-8

[4] D.J. Farace and J. Frantzen, ed. (1998), GL'97 Conference Proceedings : Third International Conference on Grey Literature : Perspectives on the design and transfer of scientific and technical information, Luxembourg DGXIII, November 13-14, 1997. - Grey Literature Network Service. - Amsterdam : GreyNet/TransAtlantic, March 1998. p. iii. - (GL-conference series, Conference proceedings, ISSN 1386-2316 ; No. 3). –
ISBN 90-74854-17-6

[5] Grey Literature Network Service http://www.greynet.org/

[6] Op cit #2, p. 48

[7] J. Schöpfel et al. (2005), Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. – In: The Grey Journal: An International Journal on Grey Literature, TGJ Volume 1, Number 1, 2005, p. 37. – ISSN 1574-1796

[8] The Directory of Open Access Repositories – OpenDOAR http://www.opendoar.org/find.php?search=unpublished

[9] Blackboard Academic Suite https://uno.blackboard.com/

[10] Library, Information Science & Technology Abstracts (LISTA) database http://www.ebscohost.com/

Photo © Image Source

Finding EU publications has never been so easy

EU Bookshop

In EU Bookshop, you can order books and multimedia products covering a wide range of topics, from general information to specific subjects. Furthermore, EU Bookshop allows free download of almost all the publications published by the European Union. You can navigate in all EU official languages.

EU Bookshop is managed by the Publications Office

Publications Office
Publications.europa.eu

GL9 Conference Welcome

Koen Helsen
Chairman of the Antwerp Convention Bureau and Provincial Councilman, Belgium

Ladies and gentlemen,

It is a pleasure and an honor for me to welcome some 60 delegates from 18 countries worldwide to the Province and to the town of Antwerp.

A pleasure because I am the chairman of the Antwerp Convention bureau which is a free service association promoting our province for international and national conferences and for seminars and incentives.

Being in the heart of Europe we hope that your travel arrangements were easy and we want to show you in the next days the best we can and the variety of culture and leisure possibilities we have to offer;

It is an honor indeed to have in this House of the Province an international congress on a topic which is dear to all of us who work in municipalities. I am aware of your issues, since I have been a mayor and now a council member for the province.

The Province of Antwerp is complex but blooming. The province consists of 70 towns and is an area of almost 1/10th of the country. Because of our location and our port we are a major player in the middle of the most important western economic area in the heart of the European Union. With 568 inhabitants per square meter we are certainly one of the more densely populated areas in the world.

This congress, dear delegates, is extremely interesting and is a challenge. The themes that were discussed today and those that will be discussed tomorrow, the high level of the speakers and the audience, as well as the excellent organisation and support of all who have contributed to its success.

I realize your subject matter is complex. I see it in the different government departments dealing with grey literature in this very building. It has re-opened my eyes.

I clearly see that a conference on Grey literature that seeks to map the infrastructure in which grey literature is embedded is not a luxury. The themes I noticed on the programme should give sufficient possibilities for you to go home with new ideas and possibilities to develop your interests.

This brings me to what I consider very important in international congresses like this one: the opportunities it provides for intense networking. I consider networking to be the key to success in the meeting sector.

However, I would also like to invite you - in particular our foreign guests - to dwell not only on the intellectual level of this meeting but also to enjoy the culinary, cultural and social facilities of the city of Antwerp - taste, visit, and experience Antwerp and Belgium outside these council walls. This will no doubt enhance the image you have of Flanders - a vivid, active, and enterprising region; and, the organisers of this congress offer you plenty of opportunities to this end.

Ladies and gentlemen, May I wish you all a very pleasant stay in our province and an excellent Ninth International conference on Grey Literature.
I attach special importance to initiatives that aim at developing a future vision with a sufficiently critical sense. Hopefully this congress will also obtain that aim.

Science Links Japan

Gateway to Japan's Scientific and Technical Information and Grey Literature

Hiroshi Tsuda
Japan Science and Technology Agency, JST Paris Office

Science Links Japan was launched in June 2006 by Japan Science and Technology Agency (JST), a core organization for implementing Japan's science and technology policy in line with the objectives of the Science and Technology Basic Plan.

JST is the major S&T information dissemination center of Japan.

JST activities are:

1. To build and maintain bibliographic and factual databases to construct the foundation of S&T activities.
2. To make and disseminate databases of resources on researchers, research institutions and research resources.
3. To aid the digitization and publication of electronic journals of Japanese academic societies.

In this connection, JST has created and manages Commercial databases called JDream-II: Document retrieval system of academic papers from throughout the world (in Japanese). JST also manages various Free-of-charge databases such as Factual databases of researchers, research institutes, patents, job information for researchers and so on.

Science Links Japan is a topically arranged directory of online information resources for science and technology in Japan. Japan's scientific and technical information (STI) scattered across or isolated on the Internet have been collected and categorized under major topics. The Website aims to provide ease of access to Japan's STI for non-Japanese researchers, policy makers and many others who need Japan's STI.

In addition to the English and Chinese site, a French language version is scheduled to be released in March 2008.

- http://sciencelinks.jp (English)
- http://sciencelinks.jp/ch/ (Chinese)
- http://sciencelinks.jp/fr/ (French) (to be released soon)

◇ **More than 1,000 sources of Japanese science and technology information introduced by category**

Sci-Tech News, Portal Sites, Study & Research in Japan, White Papers, Databases, Online Journals, PR Magazine/Newsletter, Survey/Statistics, Operating Reports, Intellectual Property, Codes and Standards, Laws & Regulations, Academic Societies, Policy-Related Organizations

◇ **Science and technology related commentaries**

Opinions ; an insight into current scientific issues from journalists or researchers
Café Terrace ; blogs by non-Japanese who have first hand experiences of life, research, and any other important aspects about Japan

◇ **J-EAST (JST English Articles of S&T Database)**

Around 1.4 million bibliographic records in English of articles in science, technology and medicine published in Japan

◇ **Basic outlines of Japan's science and technology**

Japan's S & T Overview ; S & T Organization Chart ; Japan's Listed S & T Companies ; Japan's Universities & Institutes ; Research Positions in Japan

◇ **The number of Access to *Science Links Japan***

500,000～600,000 pageview/month

We welcome your comments and suggestions via scilinks@jst.go.jp
or the following Contact Us page: http://sciencelinks.jp/content/view/274/239/

Find the Piece That Fits Your Puzzle ...

The Grey Literature Report

from

The New York Academy of Medicine

Focused on health services research and selected public health topics, the Report delivers content from over 250 non-commercial publishers on a bi-monthly basis.

Report resources are selected and indexed by information professionals, and are searchable through the Academy Library's online catalog.

Let us help you put it all together; subscribe to the Grey Literature Report today!

The New York Academy of Medicine

For more information visit our website: www.greylit.org

Or contact us at: greylithelp@nyam.org

New York Academy of Medicine's Resource Guide for Public Health Preparedness

Lea Myohanen
NYAM, The New York Academy of Medicine, United States

http://www.phpreparedness.info

RGPHP

Resource Guide for Public Health Preparedness

The Resource Guide for Public Health Preparedness (RGPHP) had its origins in the September 11, 2001 terrorist attacks in the United States and the subsequent anthrax scare later on that year. It became clear that there was no one place where disaster related material could be located but multiple sources had to be consulted. Since 2002 the New York Academy of Medicine has worked under a grant from the National Library of Medicine identifying freely available disaster related material on the internet and disseminating links to this material through its database RGPHP at http://www.phpreparedness.info/.

Material is collected by the Project Librarian who enters it into the RGPHP database using over 200 subject headings in 4 disaster phases for 9 audience types. Material becomes available only after the Project Director authorizes the release. To date the database contains over 2000 items dealing with both man-made and natural disasters. RSS feed on the latest additions is pushed out once a week, and once a month a reminder is sent to about 100 subscribers. Subscription to RGPHP is free.

The future directions may include mirroring the database as well as transforming the database into an OAI-PMH repository. Also, convening an advisory group to discuss future directions, including annotation of resources has been discussed.

CERIF: A format to enable interoperability of research information

Brigitte Jörg
euroCRIS, Germany

Current Research Information Systems (CRISs) are intended for the management of research related information. CRISs are in operation across organisations that are involved in research activities to capture important metadata and to ensure for the performance of research related processes. In practise, research related organisations often run CRISs that do not interoperate with other CRISs, applications or repositories. Such a lack of communication not only results in islands of information that are difficult to access but also results in a loss of information quality due to redundancy on the one hand and incompleteness on the other. To overcome the information islands and to connect them towards a valuable knowledge infrastructure, we propose for a standard layer, CERIF: Common European Research Information Format.

The European Commission recommended CERIF to member states as a standard for recording research information and handed over the responsibility for CERIF to euroCRIS (http://www.eurocris.org/).

CERIF has been developed and extended since first implemented in 1991. CERIF captures research actors like people and organisations, their core research activities in projects, publications and a wider research environment such as funding programme, events, patents, products, equipment, service, etc. Not only research entities as such are represented, but CERIF also allows for a flexible and scalable capturing of the relations between those entities.

With the latest CERIF2006 release major improvements have been implemented for the management and for the application of these interrelations as a so called Semantic Layer. The semantic layer allows for a simple definition of multiple role and type schemes and supports the integration of terminologies, ontologies or other classification schemes and a mapping between them. Additionally, CERIF2006 provides an XML based data exchange specification and validation schemes.

CERIF-based systems are running in various organisations and in a wide range of research environments across European member states and beyond. We consider CERIF to be well suited not only for building quality CRISs but also for enabling communication between CRISs, applications and repositories in a European research infrastructure. Existing repositories, including those of grey literature usually lack quality metadata. CRISs provide the data that can be used as metadata to describe objects in a grey literature repository including their full research context, and their provenance. Furthermore, CERIF provides the ability to link objects in a grey literature repository to – for example – repositories of research datasets.

PUMA & MetaPub: Open Access to Italian CNR repositories in the Perspective of the European Digital Repository Infrastructure

Stefania Biagioni, Carlo Carlesi,
Giuseppe A. Romano, Silvia Giannini
CNR-ISTI Pisa, Italy

Roberta Maggi
CNR Library Genoa, Italy

The DRIVER project is realizing an infrastructure through which any form of scientific-content resource, including scientific/technical reports, research articles, experimental or observational data, rich media and other digital objects will be freely accessible, thus responding to the Open Archive Initiative vision. The project is funded by the European Commission under the auspices of the "Research Infrastructure" unit and it is building upon existing institutional repositories and networks (www.driver-repository.eu).

The Publication Management System (PUMA) <http://puma.isti.cnr.it> we are presenting here is the first step towards creating an Italian network of institutional repositories, looking at the DRIVER vision.

PUMA is a software infrastructure, user focussed and service oriented, developed by the Institute of Information Science and Technologies (ISTI-CNR). It presently manages 22 CNR institutional repositories that contain about 5200 documents covering different disciplines. Repositories and collections are growing daily.

The system functionalities are oriented to meet the requirements of CNR researchers by facilitating their self-archiving, ensuring the preservation of their documents, providing world wide easy web access to their papers and by permitting them to manipulate the stored content to fulfill scientific and administrative issues.

PUMA offers a common user interface, both in English and in Italian, to Search, Browse and Self-archiving. Authors can submit document metadata (qualified DC + administrative metadata) and document full text; they are also enabled to associate access rights to the text of their documents and are responsible for their dissemination. PUMA let administrators and librarians and/or reviewers to perform technical, documentary and quality control before documents are included into the repository. The access to the various administration services is controlled, at the different levels, by LDAP server or IP address and by user name and password.

Special functions are provided by the additional software module MetaPub that uses repository content to deliver different bibliographic and administrative services. MetaPub also opens PUMA repositories to the OAI world by implementing the OAI-PMH protocol (harvesting is allowed under authorization).

APPENDICES

AUTHOR INFORMATION

Asserson, A. **p. 9**
Anne Asserson holds a Cand. polit. with a Masters in Information Science from the University of Bergen, UiB. She has been working with Research Documentation, and has participated in substantial parts of CRIS developmental work, locally and nationally. Anne Asserson has been part of the establishing and implementing of a Research Documentation system, Fdok http://www.ub.uib.no/fdok/sok/, at the UiB. For several years she was the chairwoman of the Steering Group of the national CRIS system and project secretary of a National system for academic administration. Anne Asserson is presently representing UiB in the national group that is implementing a new national research documentation system, FRIDA. She has also participated in The CORDIS funded European-wide project on " Best Practice" 1996 . She was a member of the working group set up 1997 that produced the report CERIF2000 Guidelines (1999) www.cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model www.eurocris.org. Anne Asserson is a member of the Best Practice Task Group.
Email: anne.asserson@fa.uib.no

Biagioni, S. **p. 145**
Stefania Biagioni is a member of the technical staff at the Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI), an institute within the Italian National Research Council (CNR) located in Pisa. Her affiliations are within the Library (BIB) and Networked Multimedia Information Systems (NMIS). She has coauthored a number of publications dealing with digital libraries and is a member of GreyNet since 2005. Email: stefania.biagioni@isti.cnr.it

Blaaij, C. de **p. 74**
Cees de Blaaij studied Social and Economic History at the University of Nijmegen and Library Science at the University of Amsterdam. He worked for Ernst & Young, management consultants, and the Institute for Information Law (University Amsterdam) as information professional. At the moment he is working as an academic librarian and coordinator digital services for the Public and Academic Library of Zealand, Netherlands. He took part in several GL conferences. He published several articles on issues concerning copyright in the digital environment and accessibility of grey literature on the Internet. Email: cdbl@zebi.nl

Castriotta, M. **p. 31**
Maria Castriotta is a Documentalist Researcher at the National Italian Institute for Safety and Health at Work since 1995. She is involved in studies regarding information systems related to Occupational Safety and Health, with the use of various supports. She has promoted, among others, two projects for the construction of the ISPESL Grey Literature Databank and of the Research Databank. She participated in several workgroups within the European Agency for Safety and Health at Work in Bilbao and is involved in the Topic Centre Work and Health for the maintenance of the Agency Webpage on Research. She formerly worked for 14 years at the National Italian Institute of Health, using online systems for biomedical documentation retrieval. Email: m.castriotta@doc.ale.ispesl.it

Di Cesare, R. **p. 39**
Rosa Di Cesare was born in Civita d'Antino (AQ) in 1952 and graduated from "La Sapienza" University in Rome in 1982. She received her diploma in Librarianship from the Vatican Library in 1996. She worked in the Central Library of National research council (CNR), where she started to become involved in research activity in the field of Grey literature (GL). Member of the Technical Committee for the SIGLE database from 1995 to 2001, she is presently responsible for the Library at the Institute of research on population and social policies (IRPPS) of the National research council. Her studies have focused on citation analysis and on the use of GL in scientific publications. Email: biblio.irpps@irpps.cnr.it

Dijk, E. **p. 15**
Elly Dijk is a graduate of the University of Amsterdam, where she studied Human Geography and Documentary Information Science. She is an editor at the Department of Research Information of the Royal Netherlands Academy of Arts and Sciences (KNAW). She has been involved in the development of the Dutch national research database, CRIS (Current Research Information System). She is also the coordinator of DAREnet, the portal developed as part of the Digital Academic REpositories (DARE) programme. This is a joint initiative by the Dutch universities to make their research results available in digital form. In addition, she is a member of the DARE & Co (Communication) working group of the DARE programme (2003-2006) and its follow-up programme SURFshare (2007-2010). Elly Dijk also is member of the project group for NARCIS, a portal which integrates the scientific information from CRIS, repositories and archives with datasets. She is a member of the programme committee of CRIS2008, to be held in Maribor, Slovenia, in June 2008. Email: elly.dijk@bureau.knaw.nl

AUTHOR INFORMATION
(continued)

Farace, D.J. **p. 133**
Dominic J. Farace is Director of TextRelease, an Amsterdam based information bureau specializing in grey literature and networked information. He is a native Louisianan and holds two degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked for the past twenty-seven years. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (Swidoc/KNAW), he 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; Guest Editor for PRQ, Publishing Research Quarterly and Editor of TGJ, The Grey Journal. Email: info@textrelease.com

Frantzen, J. **p. 133**
Jerry Frantzen graduated in 1999 from the College of Amsterdam in Library and Information Science. He is presently employed with Boer & Croon, Strategy and Management Group in Amsterdam. Frantzen is the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet, Grey Literature Network Service, as a freelance technical consultant.

Gelfand, J. **p. 113**
Julia Gelfand has been a librarian with the University of California, Irvine Libraries since 1981. She has been tracking the grey literature movement since the late 1980s and has participated in all of the previous GL conferences and has published and presented widely on different topics in grey literature. Her particular interests are in scholarly communications, electronic publishing, collection development, bibliography of science and technology, and she thinks that with more emphasis on networking and digital libraries, Grey Literature has a very interesting future. She is currently the chair of the IFLA Science & Technology Section and vice-chair/chair-elect of the ALA ACRL Science & Technology Section. Email: jgelfand@uci.edu

Gentil-Beccot, A. **p. 107**
Anne Gentil-Beccot studied French literature before obtaining a master degree in Information Science. She is now serials librarian in the CERN Scientific Information Service and contributes to the CERN Open Access actions.

Jeffery, K.G. **p. 9**
Keith Jeffery is currently Director, IT and International Strategy of STFC (Science and Technology Facilities Council), based at Rutherford Appleton Laboratory in UK. Previously he was Head of Business and Information Technology Department with a staff of 140 supporting over 360000 users, developing software for business and science and doing leading edge R&D. STFC hosts the UK and Ireland Office of W3C and develops and supports the largest OA (Open Access) institutional repository in UK. Keith is a Fellow of both the Geological Society of London and the British Computer Society. He is a Chartered Engineer. He is an Honorary Fellow of the Irish Computer Society. He is president of euroCRIS (eurocris.org) and of ERCIM (ercim.org) and holds three honorary professorships. He has extensive publications and has served on numerous programme committees and research grant review panels. He has particular interests in 'the research process' and the relationship of hypotheses, experiments, primary data and publications based on research in information systems, knowledge-based systems and metadata. Email: k.g.jeffery@rl.ac.uk

Jörg, B. **p. 144**
Brigitte Jörg studied information science, information systems and business administration at Saarland University, where she graduated by receiving a Magister (M.A.). She works as a researcher in the Language Technology Lab at the German Research Center for Artificial Intelligence (DFKI). Since 2001 Brigitte was involved with several development cycles of the LT World portal (http://www.lt-world.org/), architecture and system as well as with the maintenance and updates of content in the wider range of Language Technology. Since early 2005 Brigitte is manager of the European IST World project (http://www.ist-world.org/), with partners from 15 European countries, to integrate and analyze European research information in IST. Since 2004 she has been a member of the CERIF task group at euroCRIS and was appointed CERIF task group leader in early 2007. Email: brigitte.joerg@dfki.de

Luzi, D. **p. 31, 39**
Daniela Luzi is researcher of the National Research Council at the Institute of research on populations and social politics. Her interest in Grey Literature started at the Italian national reference centre for SIGLE at the beginning of her career and continued carrying out research on GL databases, electronic information and open archives. She has always attended the International GL conferences and in 2000 she obtained an award for outstanding achievement in the field of grey literature by the Literati Club. Email: d.luzi@irpps.cnr.it

AUTHOR INFORMATION

(continued)

Myohanen, L. **p. 143**
Lea Myohanen (BS, MS, MSILS) has worked as a reference librarian at the New York Academy of Medicine library since 1999. She is the library liaison to the Divisions of Health & Science Policy and Public Health and to the Center for Urban Epidemiological Studies. lmyohanen@nyam.org

Pavlov, L.P. **p. 85**
Leonid 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 GL. pavlov@vntic.org.ru

Rabina, D.L. **p. 127**
Debbie Rabina is a Assistant Professor at Pratt Institute, School of Information and Library Science. Her areas of teaching and research include scholarly communication, LIS education, government and NGO information sources, and information policy. She can be reached at drabina@pratt.edu

Schaer, P. **p. 23**
Philipp Schaer works as a researcher at GESIS-IZ Social Science Information Centre in Bonn and holds a degree in computer science from the University of Koblenz-Landau. Since 2007 he is working on the project SSOAR, funded by Deutsche Forschungsgemeinschaft (DFG), which has the goal to build an international Social Science Open Access Repository. His research interests are in human-computer interaction and user-centered design. philipp.schaer@gesis.org

Schöpfel, J. **p. 65, 133**
Joachim Schöpfel obtained his Ph.D. in psychology from the Hamburg University in 1992. During his studies in psychology, he participated in research on bilingual children of Turkish immigrants in Hamburg, of the German minority in Denmark, and in a French-German High School in Versailles, France. At present he is head of the e-publishing and document supply department at INIST-CNRS and lecturer at the Nancy University. He published on GL, document delivery, digital libraries and usage statistics. schopfel@inist.fr

Shen, W. **p. 23**
Wei Shen has a degree in computer science from Technical University Darmstadt and works since 2003 as a researcher at GESIS-IZ Social Science Information Centre in Bonn. She was involved in several projects funded by Deutsche Forschungsgemeinschaft (DFG) to enhance the technical infrastructure for scholarly publishing. In one of these projects she developed a module for in-context discussion in electronic publications. Her research interests include human-computer interaction information systems.
Email: wei.shen@gesis.org

Stempfhuber, M. **p. 23**
Maximilian Stempfhuber holds a degree in information science and business computing (M. A.) from the University of Regensburg and a PhD in computer science from the University of Koblenz-Landau. He works at GESIS-IZ Social Science Information Centre in Bonn since 1995, first as a researcher in the project ELVIRA, funded by the German Federal Ministry of Economics and Technology (BMWi) where he developed an integrated information system for market research. Since 2000 he is Head of Research and Development and since 2004 Deputy Directors of GESIS-IZ. Sine 2002 he is lecturing in human-computer interaction and computational visualistics at the University of Koblenz-Landau. He is board member of several national and international organizations aiming at improving the exchange of research information, supporting scholarly publishing and promoting Open Access. His research interests are in human-computer interaction, computational visualistics, information retrieval, and digital libraries.
Email: max.stempfhuber@gesis.org

Stock, C. **p. 39, 133**
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).
E-Mail: christiane.stock@inist.fr

Tsuda, H. **140**
Hiroshi Tsuda is Director of Paris office of Japan Science and Technology Agency (JST). He joined Research Development Corporation of Japan (JRDC), the forerunner of JST, in 1992. Since then, he has been involved with a variety of JST activities such as basic research programs, technology transfer programs and research exchange programs. He also temporarily worked for the Science and Technology Agency, Prime Minister's Office as a special staff to coordinate APEC activities and to manage international collaborative research programs as well as fellowship programs. Since 2005, he has been working in Paris to promote international collaboration between JST and counterpart organizations in Europe. He holds a bachelor of law (Japan) and diploma in business administration (US).

LIST OF PARTICIPATING ORGANIZATIONS

Antwerp Convention Center	Belgium
Belgian Health Care Knowledge Centre, KCE	Belgium
British Library, BL	United Kingdom
Centre National de Recherche Scientifique, CNRS	France
Centre of Information Technologies and Systems, CITIS	Russia
Community Research & Development Information Service, CORDIS	Luxembourg
Dartmouth College; I3P	United States
Department of Economy, Science and Innovation, EWI	Belgium
Department of Foreign Affairs	South Africa
Flemish Government	Belgium
EBSCO Information Services	United States
EU Bookshop	Luxembourg
euroCRIS, Current Research Information Systems	Netherlands
European Organization for Nuclear Research, CERN	Switzerland
German Research Center for Artificial Intelligence, DFKI	Germany
GESIS Service Agency Eastern Europe	Germany
Grey Literature Network Service, GreyNet	Netherlands
Information International Associates, IIa	United States
Information Today, Inc.	United States
Institut de l'Information Scientifique et Technique, INIST	France
Institute of Information Science and Technologies, ISTI	Italy
Institute of Research on Population and Social Policies, IRPPS	Italy
International Atomic Energy Agency, IAEA	Austria
International Centre for Informatics and Electronics	Russia
International Federation of Library Associations and Institutions, IFLA	Netherlands
Jagiellonian University	Poland
Japan Atomic Energy Agency, JAEA	Japan
Japan Science and Technology Agency, JST	Japan
Library of Zeeland, ZEBI	Netherlands
Link s.r.l.	Italy
Loughborough University	United Kingdom
National Institute of Occupational Safety and Prevention, ISPESL	Italy
National Research Council, CNR	Italy
New York Academy of Medicine, NYAM	United States
Office of Scientific and Technical Information, OSTI	United States
Official Publications of the European Communities, OPOCE	Luxembourg
Osrodek Przetwarzania Informacji, OPI	Poland
Pratt Institute, School of Information and Library Science	United States
Royal Netherlands Academy of Arts and Sciences, KNAW	Netherlands
Science and Technology Facilities Council, STFC	United Kingdom
Scientific and Technical Information Center, VNTIC	Russia
SNDT Women's University	India
Social Science Information Centre, IZ	Germany
University of Bergen,	Norway
University of California, Irvine Libraries, UCI	United States
University of Ljubljana	Slovenia
University of Salamanca	Spain
University of South Florida, USF	United States
U.S. Department of Energy, DOE	United States
Zentral- und Landesbibliothek Berlin, ZLB	Germany

GreyNet Members
20% Discount

Grey Foundations in Information Landscape

Publication Order Form

NINTH INTERNATIONAL CONFERENCE ON GREY LITERATURE *Antwerp, Belgium, 10-11 December 2007*

Forthcoming Publications (January 2008)	*No. of Copies*	x	*Amount in Euros*	Subtotal
GL9 CONFERENCE PROCEEDINGS - Printed Edition ISBN 978-90-77484-10-4 ISSN 1386-2316 *Postage and Handling excluded*[*)]		x	80.00 =	€
GL9 CONFERENCE PROCEEDINGS - CD-Rom Edition ISBN 978-90-77484-10-4 ISSN 1386-2316 *Postage and Handling included*		x	80.00 =	€
EXCLUSIVE OFFER: GL9 Conference Proceedings - CD-Rom Edition With accompanying PowerPoint Presentations *Postage and Handling included*		x	99.00 =	€
POSTAGE AND HANDLING PER PRINTED COPY[*)]				
Holland		x	5.00	€
Europe		x	10.00	€
Other		x	20.00	€
			TOTAL	€

Customer information

Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
E-mail Address:	

Check one of the boxes below for your Method of Payment:

- ☐ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
  BIC: RABONL2U IBAN: NL70 RABO 0313 5853 42, with reference to "GL9 Publication Order"
- ☐ Bank check/draft made payable to TextRelease, with reference to "GL9 Publication Order"
- ☐ MasterCard/Eurocard ☐ Visa card ☐ American Express

Card No. _____ Expiration Date: _____

Print the name that appears on the credit card, here _____

Signature: _____ CVC II code: _____ *(Last 3 digits on signature side of card)*

Place: _____ Date: _____

NOTE On receipt of payment, an invoice marked paid will be sent to your postal address. If you require a *pro forma* invoice, please notify:

TextRelease
www.textrelease.com

GL9 Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: info@textrelease.com

INDEX TO AUTHORS