https://doi.org/10.26069/greynet-2026-000.604-tgj
ISSN 1574-1796

An International Journal on

Grey Literature

Summer 2008 – TGJ Volume 4, Number 2

'Access to Grey in a Web Environment'

www.textrelease.com GreyNet Grey Literature Network Service www.greynet.org

The Grey Journal
An International Journal on Grey Literature

COLOPHON

CIP
The Grey Journal (TGJ): An international journal on grey literature / D.J. Farace (Journal Editor); J. Frantzen (Technical Editor) ; GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 4, Number 2, summer 2008. - *The British Library, INIST-CNRS, and NYAM are corporate authors and associate members of GreyNet. This serial publication appears three times a year - in spring, summer, and autumn. Each issue is thematic and deals with one or more related topics in the field of grey literature. The Grey Journal appears both in print and electronic formats.*
ISSN 1574-1796 (Print)
ISSN 1574-180X (E-Print/PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:
€110 individual, including postage & handling
€225 institutional, including postage & handling

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

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

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

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

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

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

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

Contents
"ACCESS TO GREY IN A WEB ENVIRONMENT"

https://doi.org/10.26069/greynet-2026-000.604-tgj

EDITOR'S NOTE

In this issue of TGJ, a number of articles dealing with access to grey literature in a web environment have been brought together. These articles are of European origin and reflect various approaches to this topic as well as various stages in which open access is realized. Since its relaunch in 2003, GreyNet has actively followed the developments in the OAi movement and recently signed on to the OpenSIGLE Repository http://opensigle.inist.fr/. In so doing, GreyNet seeks to better preserve and make openly available research results originating in the International Conference Series on Grey Literature. Moreover, the choice of OpenSIGLE will lead LIS professionals to the crossroads of more than 25 years of bibliographic information on grey literature with 15 years of research in this field of study.

GreyNet together with colleagues at INIST-CNRS designed the format for a metadata record, which encompasses standardized PDF attachments of the authors' conference preprints, PowerPoint presentations, abstracts and biographical notes. The first test batch containing records from the Eighth International Conference on Grey Literature (GL8, 2006) was uploaded. A few minor problems that were encountered have since been successfully resolved and these records together with those from GL7 (2005) are now available for search and retrieval in OpenSIGLE. By December 2008, conference records over the past five years including those from GL9 (2007) will be housed in the OpenSIGLE Repository. Those records from the earlier four conferences in the GL Series (1993-1999) will require additional image scanning as well as permission from Emerald (the former MCB University Press). Should this be granted, then the total number of GreyNet records in OpenSIGLE would be doubled and the collection would become comprehensive.

Dominic J. Farace, Journal Editor
journal@greynet.org

https://doi.org/10.26069/greynet-2026-000.604-tgj

Accessing grey literature in an integrated environment of scientific research information*

Elly Dijk, Chris Baars, Arjan Hogenaar, and Marga van Meel
(The 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.

* First published in the GL9 Conference Proceedings, February 2008

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)[1] 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.

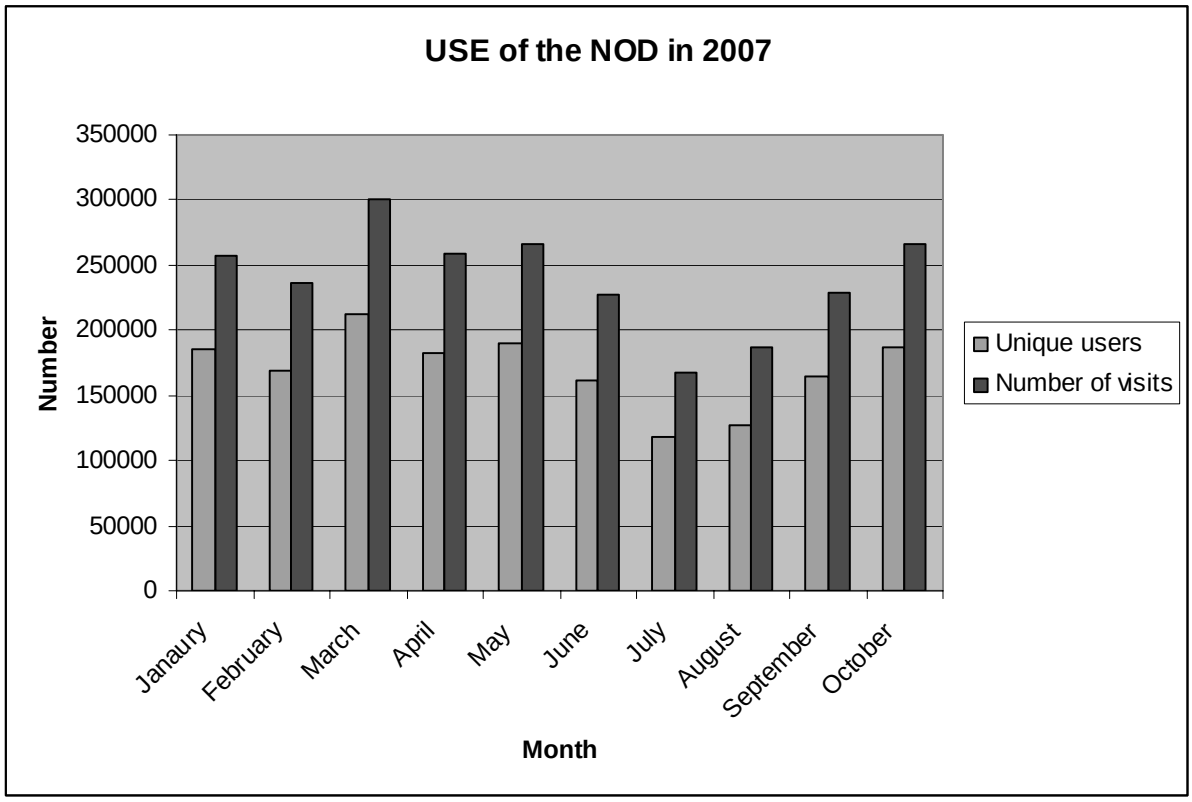


Figure 1. Use of the NOD in 2007

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[2] website gives access to these repositories. This is referred to as the so-called "green route of Open Access publishing".[3] 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).

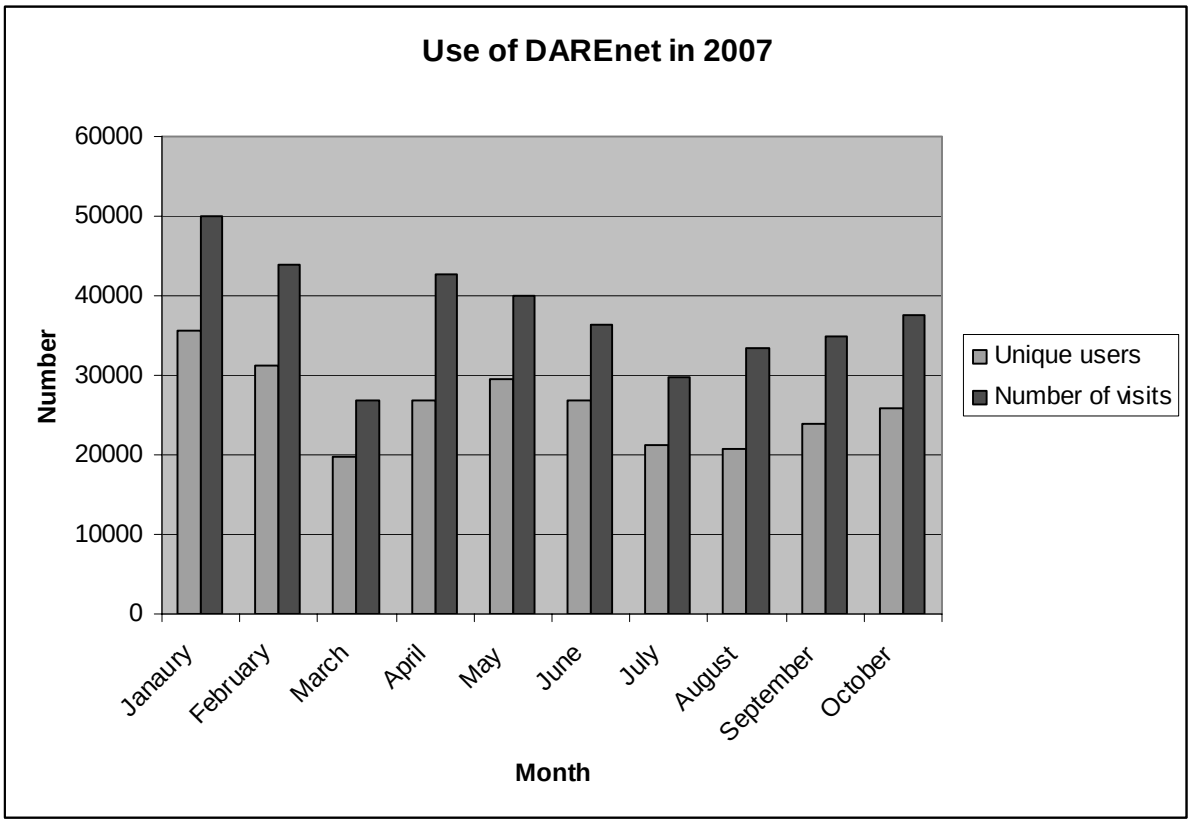


Figure 2. Use of DAREnet in 2007

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[4]. 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).

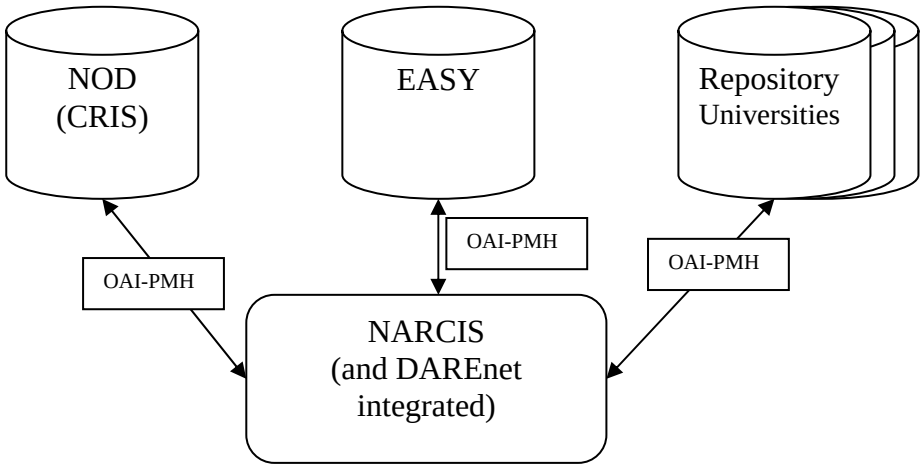


Figure 3. NARCIS harvest the national CRIS, EASY and the academic repositories

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.[5] 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)"[6], 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[7].

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[8]). 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.

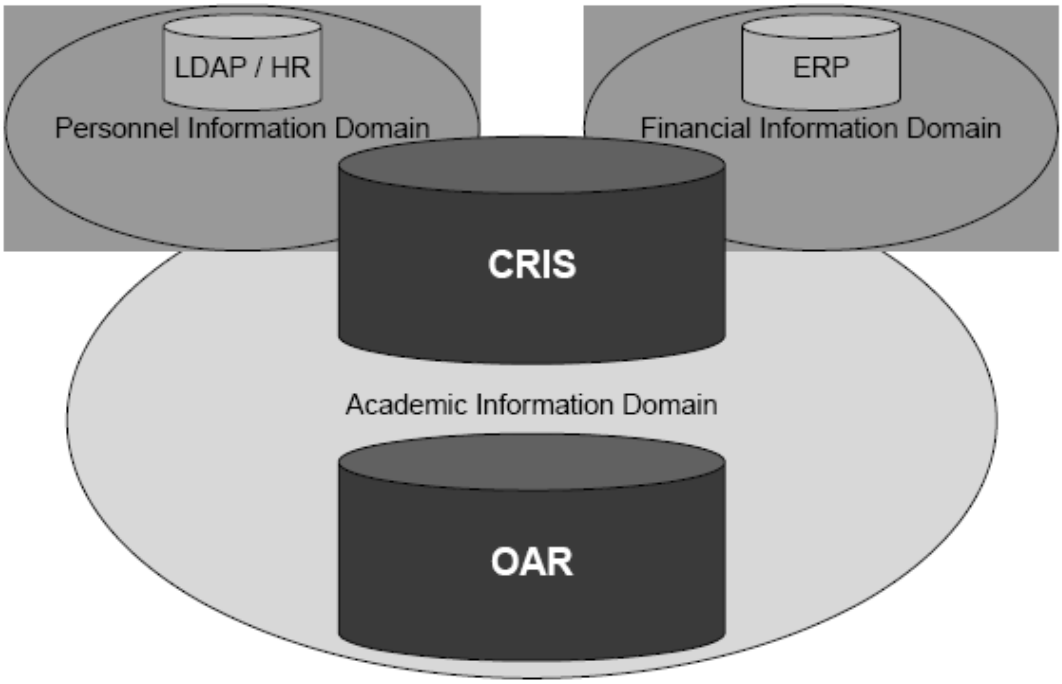


Figure 5: the Academic Information Domain

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)[9] 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).[10] 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)[11]. 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'[12].

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

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

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

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

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

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

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

[7] 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

[8] 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

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

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

[11] 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)

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

https://doi.org/10.26069/greynet-2026-000.604-tgj

The long and winding road: Institutional and disciplinary repository at Saarland University and State Library *

Ulrich Herb and Matthias Müller
(Germany)

Abstract

Purpose of this paper

The article describes the Open Access archiving of scientific primary and secondary literature on the institutional and the disciplinary repository of the Saarland University and State Library. The challenges, opportunities and success factors of both repositories are compared and evaluated.

Approach

The article is based on the findings and experiences of eight years of Open Access archiving practice.

Findings

Basic requirements for success and acceptance are

- quality control and standardisation; that could be proved for example by an official certificate,

- visibility and dissemination; researchers emphasize that their documents are indexed by the databases most common within their community. This applies especially to disciplinary repositories.

Value of paper

Universities should undoubtedly offer an institutional repository. Usually the university library runs this repository. The article lists some key points that should be considered when an organisation plans to start respectively run a institutional or disciplinary repository.

1.0 Introduction

The Library of Saarland University and the State of Saarland (Saarländische Universitäts- und Landesbibliothek, SULB)[1] is the central academic library of Saarland University,[2] Germany. At the same time, SULB is participating in a national program for scientific literature supply organized by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).[3] As a result of this dual role SULB maintains and runs two Open Access repositories:

- SciDok, http://scidok.sulb.uni-saarland.de, an institutional data repository for all the scientists at the University of Saarland regardless of their discipline and
- PsyDok, http://psydok.sulb.uni-saarland.de, a disciplinary data repository for psychologists from the German-speaking countries.

By comparing both repositories this article aims to provide an insight into both the challenges and the chances of institutional and disciplinary repositories in Germany.

2.0 At the beginning of the road

Pre-print servers can be considered as the beginning of Open Access publishing. Several pre-print servers established especially for the sciences have come to enjoy an excellent reputation in the academic community. As a consequence, scientists increasingly asked for the implementation of electronic publishing platforms. For the most part, university libraries offered these services because they were able to guarantee the long-term availability of this new publishing distribution outlet. The prevailing software for digital repositories in Germany is OPUS because it is tailored to the special needs and requirements of the German e-publishing community.

OPUS[4] was developed at the University of Stuttgart (Germany). Meanwhile, several universities are collaborating in the further development of OPUS; SULB has also contributed several modules. To date, more than 50 German universities are using OPUS to run their digital repositories. The systems requirements include UNIX/Linux

* First published in CLC Systems & Services, Issue 1, Volume 24, 2008

with a MySQL- or MSQL-Database and an APACHE/PHP-environment. OPUS is also mentioned in the Open Society Institute's guide to institutional repository software (Open Society Institute, 2004). The preferred file formats to be published are PDF-, PostScript- or HTML-files, the ideal format to date is PDF/A.
The repositories offer a variety of retrieval options such as

- browsing through document types, faculties and institutions as well as subject classifications (e.g. DDC, Computing Classification System (CCS), Mathematics Subject Classification (MSC), Physics and Astronomy Classification Scheme (PACS) on SciDok and APA Classification and Indexing System on PsyDok)
- meta-data search
- fulltext search (PsyDok even offers a combined meta-data/ fulltext search via the SCIRUS search engine)

In addition to this, users can subscribe to an alerting service which is realised as an RSS feed.

2.1 Open Access and long term accessibility and data preservation

Both SciDok and PsyDok are Open Access servers, which means publishing and document use are free of charge. Many different document types may be published and found in these repositories: diploma and master theses, dissertations, professorial dissertations, preprints, teaching materials, psychological scales, grey literature, monographs, book chapters, journals and of course articles from journals.
The SULB Library ensures permanent and long-term access, long-term usability and **citeability** for all documents published. In order to provide a permanent and stable electronic location for its documents, both repositories offer a consistent identifier in the form of a Uniform Resource Name[5] (URN) for each document. Each URN is registered by the German National Library[6] (Die Deutsche Nationalibliothek, DNB). As a State Library with the duty of archiving electronic documents, SULB uses the results and findings of relevant German projects such as NESTOR[7] or KOPAL[8] in order to guarantee the long-term access to and availability of its repositories. Technologies and practices developed and evaluated within these projects will be applied to PsyDok and SciDok. Further features to be implemented as soon as possible include multilingual user interfaces, cross-lingual search and interoperable download statistics for each document.

2.2 Quality and standardisation of the repositories

When speaking of standardisation and technological and organizational quality control of digital publication services in Germany, the German Initiative for Network Information (German: Deutsche Initiative für Netzwerkinformation e.V., DINI)[9] has to be mentioned. DINI developed a variety of standards digital repositories have to comply with in order to become certified DINI-Servers. DINI strives to lend greater credence and reliability to electronic publishing with digital repositories; the DINI Certificate is a mark of quality. The criteria checked by DINI-experts include organizational matters (e.g. workflow optimization), a dedicated policy for the repository, questions of data integrity and authenticity, guaranteed archival storage periods, detailed description of the service, data export into long-term archival systems, questions of content indexing and question of meta-data management and export. PsyDok and SciDok meet all these criteria and are certified DINI-Servers. Another important proof of the adherence to standards is of course the compliance of these two repositories with the Protocol for Meta-data harvesting of the Open Archives Initiative.[10]

2.3.1 SciDok: An institutional repository for scientists from Saarland University

The SciDok digital repository was launched in 2001. Initially, SciDok - just like most German repositories - was used for the electronic publication of dissertations. At the beginning, the task of winning support for SciDok was fraught with obstacles: visiting scientists on-site to introduce SciDok to them, staging campaigns for electronic publication, obtaining modifications of internal university regulations governing the publications of the university's members, giving Ph. D. students the possibility to publish their dissertations in electronic form etc. At the same time, an awareness of the advantages inherent in depositing other publications (e.g. research papers) in the repositories had to be developed. This is the reason why so many dissertations can be found on German Open Access Repositories: in German repositories, grey and white literature is usually mixed.
The scientists responded to SciDok in two different and opposing ways: natural scientists, computer scientists and physicians on the one hand received the new

offering positively, while researchers from the Humanities on the other hand were hampered by a lack of confidence in the "immaterial" character of online technology.

2.3.2 PsyDok: A disciplinary repository for the psychological community

In the fall of 2002, SULB and the Institute for Psychological Information (Zentrum für Psychologische Information und Dokumentation ZPID, Germany)[11] launched a joint project named „Digital Psychology Information (DPI)" which was funded by the DFG. The aim of DPI was to improve the availability and accessibility of electronic psychological information. One of the work packages defined within DPI was the implementation of the subject-specific repository PsyDok operated by SULB. By establishing a service like PsyDok, SULB responded to the ephemeral nature of most online information and at the same time made a contribution to the availability and preservation of electronic documents. In accordance with the principles laid down for special subject collections (SSG-), PsyDok acts as a supra-regional, disciplinary repository for digital scientific documents. It is a central point for psychological fulltext publishing under Open Access policies, not only for Germany but also for other German-speaking countries like Switzerland and Austria.
From the inception of the project on, special emphasis was put on grey literature. In contrast to SciDok, where different faculties and their attitudes towards online documents have to be considered and taken into account, PsyDok has only one single clientele: the psychological community. Hence, PsyDok was advertised at psychologists' conferences and meetings and a general acceptance was easily achieved. This is the reason why PsyDok was and is more frequently used than SciDok.

3.0 The road splits and turns to green

Scientists cannot and do not want to abandon the publication of their articles in scientific journals. The main reason for this is the peer review process guaranteed by the publishers. This quality control and the attendant gain in reputation linked to this process provides an incentive for academic authors to publish preferably in such journals. Repositories – especially the multidisciplinary repositories of universities – cannot offer a comparable mechanism for quality control. As a result, the desire to collect journal articles in open access repositories in order to make them freely available arose soon after. Initially, academic publishers tried to prevent this by referring to copyright issues. However, sustained pressure on the publishers by universities, libraries, funding organsisations and, last but not least, the authors themselves is gradually effecting a change, so that it is now getting easier over time to deposit digital copies of journal articles in the repositories. SHERPA creates both public awareness and puts political pressure on the publishers by using an system of colour codes: it assigns publishers a certain colour according to their copyright and self-archiving policies. Publishers allowing parallel depositing of postprints or preprints of journal articles are rated as 'green' publishers. Hence the metaphor "green road". Meanwhile, all the important international publishers like Elsevier or Springer have become green publishers. Unfortunately, most of the smaller German publishers still have difficulties with adopting a green copyright policy.
The SULB library is aiming to get as many journal articles into SciDok as possible. The starting point is the annual bibliography ("Jahresbibliographie") of Saarland University. Here, all publications of all university members have to be collected. Initially, contributions were voluntary and the cooperation from the scientists was poor. But since the university administration is now using the figures of this bibliography as the basis for the evaluation of the faculties and departments, all researchers at Saarland University are taking pains to contribute all their publications.

A unique feature of SciDok is its dovetail with the annual research bibliography of the Saarland University:[12] Scientists can simultaneously contribute publications for the annual research bibliography and publish the corresponding documents in SciDok in one single step. If scientists register only the metadata in the annual research bibliography, SULB staff members check the publisher´s copyright policies for self-archiving in the SHERPA database.[13] If a parallel publication is possible, a SULB staff member will ask the author for the full text and publish it in the repository using the metadata already stored in the bibliography. Metadata in the annual research bibliography and full texts in SciDok are interlinked.

3.1 Dissemination and Visibility

The primary beneficiaries of these repositories are the scientists themselves. Depositing documents in Open Access repositories offers impressive advantages

especially when compared with the still customary but less-than-ideal practice of publishing documents on web servers of academic institutions or faculties. Such information cannot be searched (or retrieved) systematically; if a document is found at all, it is found accidentally by click-by-click-retrieval. In addition to this, publications like these are of a very limited value in the scientific discourse: both the life-span and availability of such documents are short (or at least unpredictable) and permanent access to these data, their immutability and quotability, are doubtful.
By offering a rapid and global flow of scientific information, PsyDok and SciDok exemplify the motto "maximizing research impact by maximizing research access". Numerous studies have proved that the additional publication of documents in an Open Access repository increases the number of its citations (Lawrence 2001, Brody et al. 2005, Brody et al. 2004, Hajjem et al. 2005, Eysenbach 2006). The number of citations also depends on the visibility of the documents; that is another incentive for authors to publish their documents in these repositories.
Unfortunately, most publishers expressly allow the depositing of a parallel version of the articles only in *university* repositories. Disciplinary repositories like PsyDok are seldom mentioned as another publication outlet in the SHERPA list. Thus, another grey area has grown as authors are uncertain whether to contribute articles to PsyDok in addition to the conventional academic publishing outlets or not. This divergence between institutional and disciplinary repositories is likely to disappear, however, because both authors and publishers stand to benefit from a parallel publication: documents deposited in Open Access repositories achieve greater impact and gain greater exposure because of the dissemination via the Open Access and scientific search and retrieval infrastructure, e.g. the Open Archive Initiatives Protocoll for Metadata Harvesting (OAI PMH).[14]
Hence, the SULB library makes every effort to integrate both repositories in disciplinary and multidisciplinary retrieval systems.

3.2.1 SciDok

The documents on SciDok are indexed by numerous search engines, databases and catalogs:

- SULB-OPAC,[15]
- Karlsruher Virtueller Katalog (Karsruhe virtual catalog, a meta search engine for library catalogs),[16]
- OPAC of the Südwestdeutscher Bibliotheksverbund (SWB),[17]
- Bielefeld Academic Search Engine (BASE),[18]
- Open access to scientific literature (OASE),[19]
- OAI based search engines,[20]
- Google Scholar,[21]
- Scirus[22]/SCOPUS[23] of Elsevier,
- Thomsons Web Citation Index,[24]
- partly indexed in subject databases (e.g. CiteSeer[25] or IOPort[26])

3.2.2 PsyDok

PsyDok offers all psychologists – regardless of their university or institution – the opportunity to make their scientific output freely available and accessible. Disciplinary repositiories are very attractive for database producers and search engine operators because the entire information in such a repository can be used. For the same reason, disciplinary repositories are also appealing for scientists: a highly-regarded repository very likely paves the way into a particular community´s well-established databases.
In addition to the information sources mentioned above, PsyDok is integrated in the following services:

- Diploma and matser theses database[27] of the ZPID,
- PsychSpider[28] of the ZPID,
- the database Psyndex[29] of the ZPID (the most comprehensive psychological database in the German speaking countries, corresponding to PsycINFO in the Anglo-American region),
- PsychLinker of ZPID and SULB,[30]
- Current Web Contents,[31]
- PsychCRAWLER[32] of the American Psychological Association (APA).

And nota bene, SCIRUS ranks PsyDok among its preferred web sources (like ArXiv or BioMed Central et al.).[33]

3.3 Issues of quality and standardisation

Usually, operators of 'green road' repositories do not have any influence on the quality of the submitted and archived documents. Hence, SULB has little chance to ensure any kind of quality control for SciDok or PsyDok. Nevertheless, quality principles can be upheld implicitly because they are bound up with the type of document. With institutional series or working paper series, an editorial review should be assumed. With dissertations and professorial dissertations a sufficient quality can be assumed because of the type of document. Diploma and master theses will only be published if the referee explicitly states a recommendation. In the case of the growing body of preprints, the authors' reputation can be based on former full publications; and the quality of a pre-print document can be judged through a comparison with the later formal publication. Postprints are controlled by the publishing houses' expert staff, mostly in the form of a peer review.
Each department or faculty can implement a peer review for their own documents.

4.0 Conclusion

The situation at Saarland University reflects the status of Open Access at most German universities. Unfortunately, the university has yet not adopted a voluntary Open Access policy, not to mention a mandatory one. At the same time, the idea of Open Access does not have a lot of currency among many scientists. This is consistent with the findings of a study conducted by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, 2005), which proves that Open Access is hardly known among German scientists. However, those scientists who became familiar with Open Access services are using them to an increasing extent.

The development of PsyDok clearly demonstrates that. After getting familiar with PsyDok, many scientists have come to appreciate the benefits of Open Access publishing, which then leads to an increasing number of publications. As a result of good experiences[34] with this publishing outlet, a number of authors have asked SULB for extension of its publication services. By now, several primary publications such as conference proceedings or monographs can be found in the two repositories.

Apart from the electronic Open Access document, authors and learned societies are asking for printed versions of those documents:

- In 2006, a theological Open Access journal[35] was founded at Saarland University and is published electronically at SULB. The editors plan to collect the articles of one full year in a printed volume.
- Several authors and institutions using PsyDok are asking to have monographs published.
- Psychologists from the University of Klagenfurt (Austria) already publish their proceedings *Beiträge zur Qualitativen Inhaltsanalyse* on PsyDok and are planning to edit a printed version additionally.

These encouraging experiences confirm SULB in its efforts to improve and expand the green road services and also to develop alternatives for golden road publishing. This means that SULB would have to take on the function of a university press.

Literature

Brody, Tim and Harnad, Stevan (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. Dlib Magazine, 10, Nr. 6, http://www.dlib.org/dlib/june04/harnad/06harnad.html

Brody, Tim, Harnad, Stevan and Carr, Les (2005). Earlier Web Usage Statistics as Predictors of Later Citation Impact. Journal of the American Association for Information Science and Technology (JASIST). http://eprints.ecs.soton.ac.uk/10713/

Eysenbach, Gunther (2006). Citation Advantage of Open Access Articles. Public Library of Science Biology 5(4), http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0040157

Hajjem, Chawki, Harnad, Stevan and Gingras, Yves (2005). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. IEEE Data Engineering Bulletin 28(4), S. 39-47 http://eprints.ecs.soton.ac.uk/11688/

Lawrence, Steve (2001). Free online availability substantially increases a paper's impact.
Nature, 411, S. 521-522
http://www.nature.com/nature/journal/v411/n6837/full/411521a0.html

Open Society Institute (2004). A guide to institutional repository software, 3rd edition.
http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf

Deutsche Forschungs-Gemeinschaft (DFG) (2005). Publikationsstrategien im Wandel? : Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten DFG-geförderter Wissenschaftler unter besonderer Berücksichtigung von Open Access / Ed. Deutsche Forschungsgemeinschaft (DFG), Weinheim, Bergstr : Wiley-VCH, 2005

References

[1] For more information about Saarland University and State Library, see http://www.sulb.uni-saarland.de
[2] For more information about Saarland University, see http://www.uni-saarland.de
[3] In Germany scientific information supply is characterized by a distributed, supra-regional system of Special Subject Collections (in German: Sondersammelgebiet, SSG) which are attended by scientific libraries. There are 121 special subject collections which can be found in more than 40 libraries all over Germany. The SSG´s main function is to build a collection of subject related scientific literature which is as complete as possible; this includes domestic and foreign research literature (including electronic media, microfilms and multimedia objects) and scientific journals. Within this system the SULB is responsible for the Special Subject Collection Psychology since 1966. See http://www.dfg.de/en/research_funding/scientific_instrumentation_and_infrastructure/scientific_library_services_and_information_systems/project_funding/funding_objectives/ for more information.
[4] For more information about OPUS, see http://opus.uni-stuttgart.de
[5] For more information about the URN, see http://www.persistent-identifier.de/?lang=en
[6] For more information about DNB, see http://www.d-nb.de
[7] For more information about nestor, see http://www.langzeitarchivierung.de/index.php?newlang=eng
[8] For more information about kopal, see http://kopal.langzeitarchivierung.de/
[9] For more information about DINI, see http://www.dini.de/
[10] http://www.openarchives.org
[11] ZPID creates electronic databases on literature, tests, audiovisual media, and Web resources. For more information, see http://www.zpid.de
[12] http://jahrbib.sulb.uni-saarland.de
[13] http://www.sherpa.ac.uk/about.html
[14] http://www.openarchives.org
[15] http://opac.sulb.uni-saarland.de
[16] http://www.ubka.uni-karlsruhe.de/kvk.html
[17] http://swb.bsz-bw.de
[18] http://base.ub.uni-bielefeld.de/
[19] http://www.ubka.uni-karlsruhe.de/kvvk.html
[20] e.g. OAISTER, http://oaister.umdl.umich.edu/o/oaister/
[21] http://scholar.google.com
[22] http://www.scirus.com
[23] http://www.scopus.com
[24] http://scientific.thomson.com/press/2005/8298416/
[25] http://citeseer.ist.psu.edu/cs
[26] http://www.io-port.net/
[27] http://www.psyndex.de/index.php?wahl=products&uwahl=frei&uuwahl=thesesinfo
[28] http://www.psychspider.de
[29] http://www.psyndex.de/
[30] http://www.psychlinker.de
[31] http://www.scientific.thomson.com/free/essays/selectionofmaterial/cwc-criteria/
[32] http://www.psychcrawler.com
[33] http://www.scirus.com/srsapp/aboutus/#sources
[34] “PsyDok can be considered as an alternative to conservative publication modalities: high visibility, long-term citability and free accessibility are the platform´s advantages.” (Mayring, Brunner, 2006)
[35] http://aps.sulb.uni-saarland.de/theologie.geschichte/

https://doi.org/10.26069/greynet-2026-000.604-tgj

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

Christiane Stock
(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

* First published in the GL9 Conference Proceedings, February 2008

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:

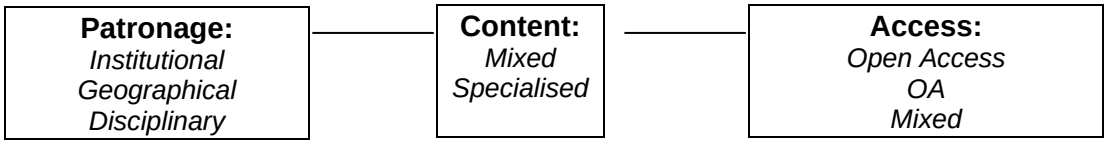


Figure 1: Basic repository classification by Andrews

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 <http://www.era.lib.ed.ac.uk> Similar academic regulations are found at many other universities like the University of Leiden (Netherlands) <http://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 Toegankelijkheid	Only internally accessible from the K.U.Leuven intranet 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.

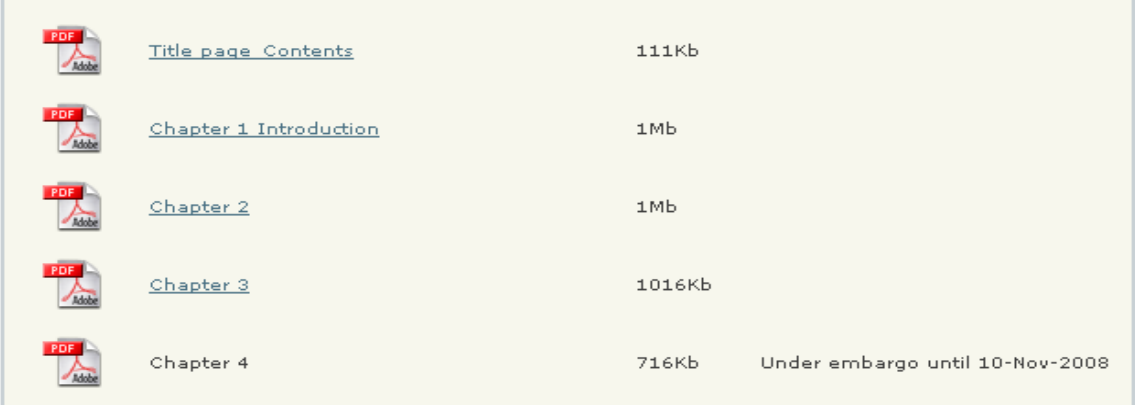


Figure 4: Leiden Repository : parts under embargo
https://openaccess.leidenuniv.nl/handle/1887/4460

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:

<u>At the start of the search :</u>

- 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) <http://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?>.

<u>At the display of the list of results:</u>

- 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.

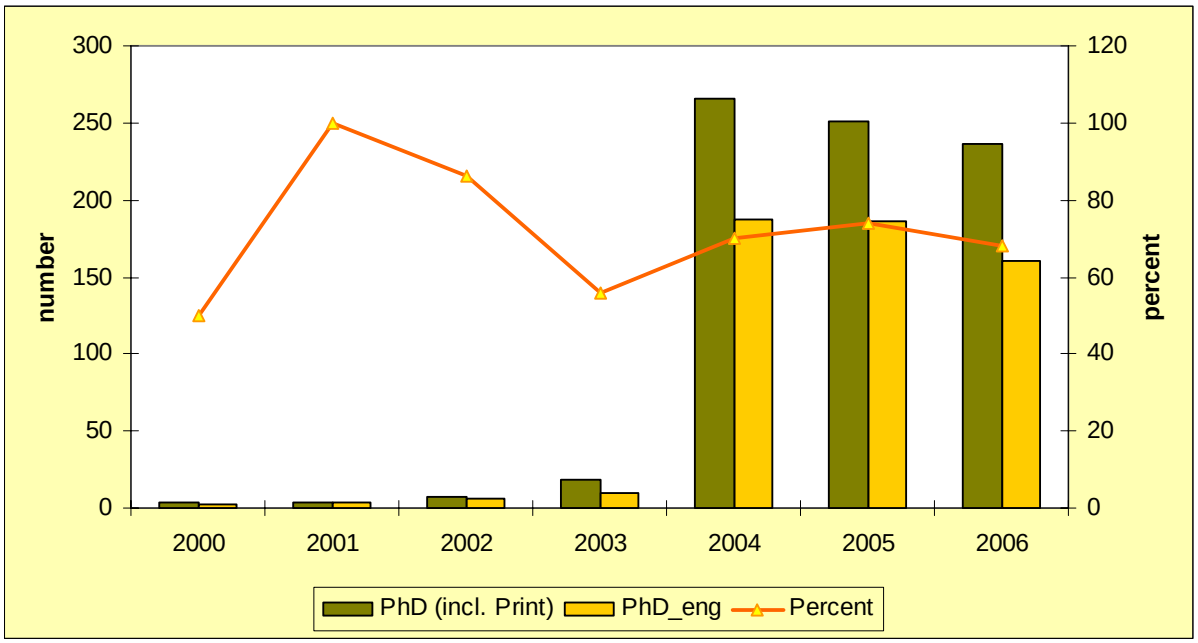


Figure 5: Stockholm University repository

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.

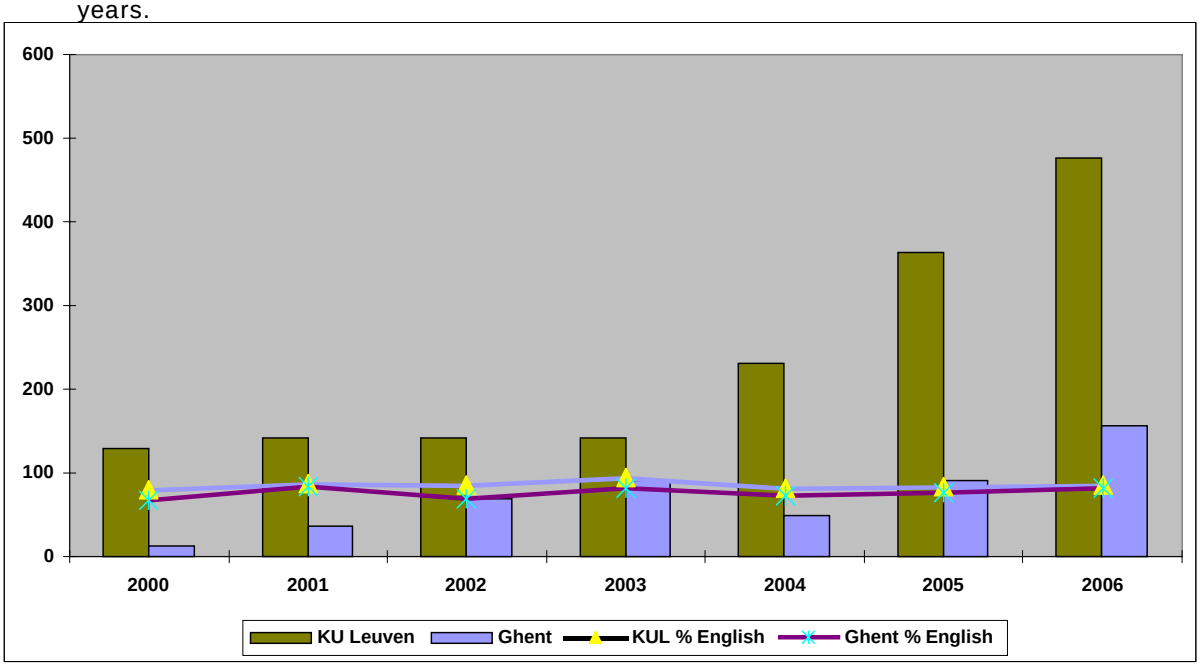


Figure 6: ETDs from KU Leuven (through LIBIS) and Ghent University

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

https://doi.org/10.26069/greynet-2026-000.604-tgj

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

Rosa Di Cesare, Daniela Luzi, and Roberta Ruggieri
(Italy)

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

* First published in the GL9 Conference Proceedings, February 2008

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

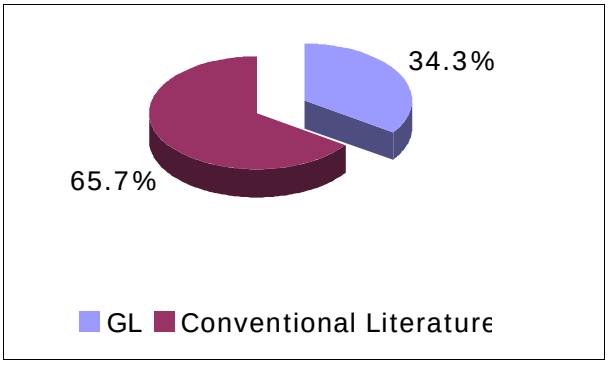


Fig. 1. – Distribution of the cited documents by publication type

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.

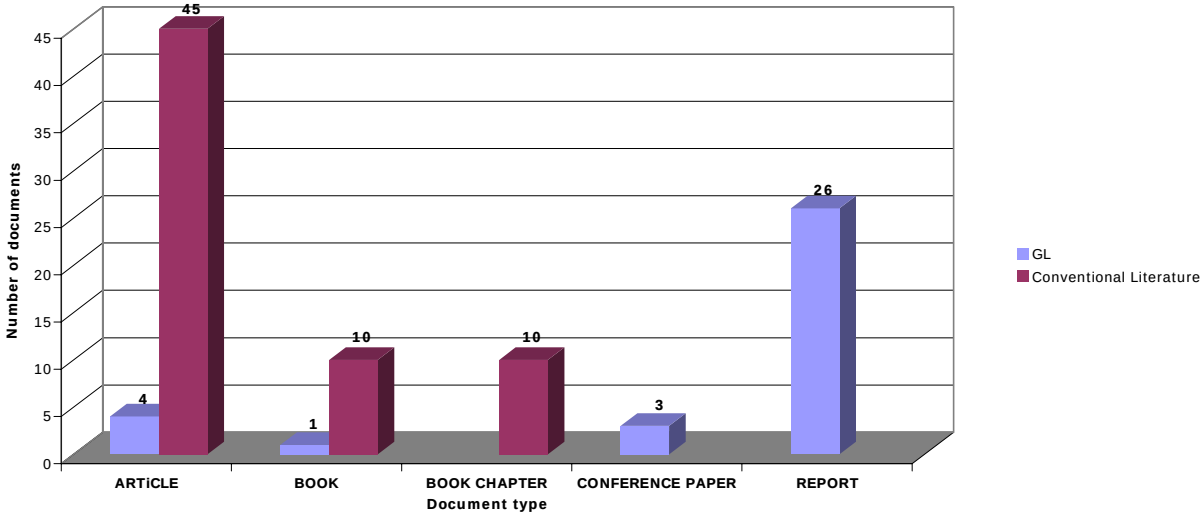


Fig. 2. – Distribution of the cited documents by document type

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.

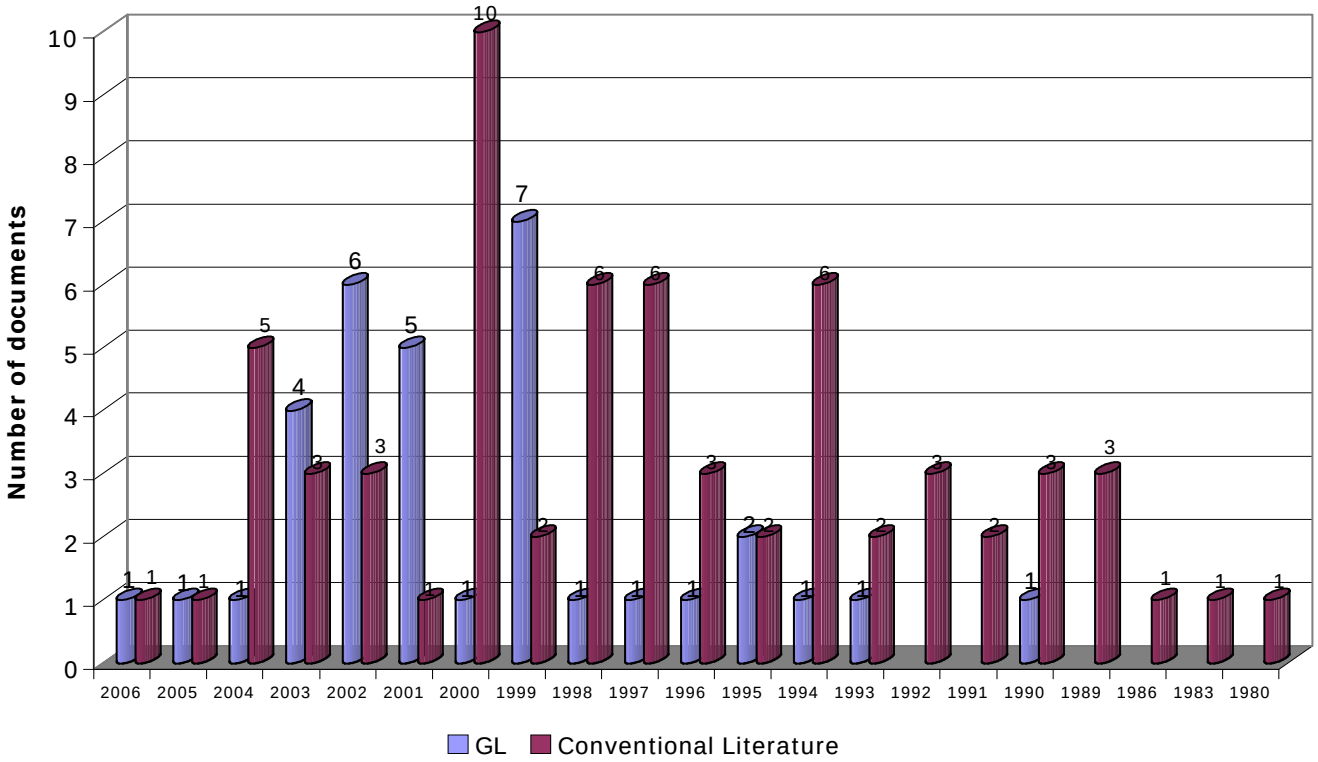


Fig. 3. - Distribution of the cited documents by publication year and publication type

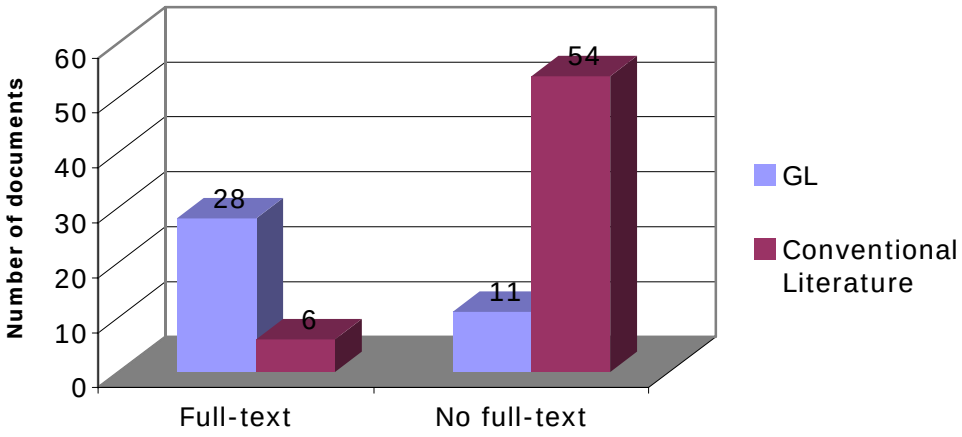
Figure 3 shows the distribution of the cited documents by publication year and publication type. The GL cited documents are published from 2006 to 1993, with the exception of one document published in 1990, while the cited conventional documents are published in a larger time span, from 2006 until 1980. 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].

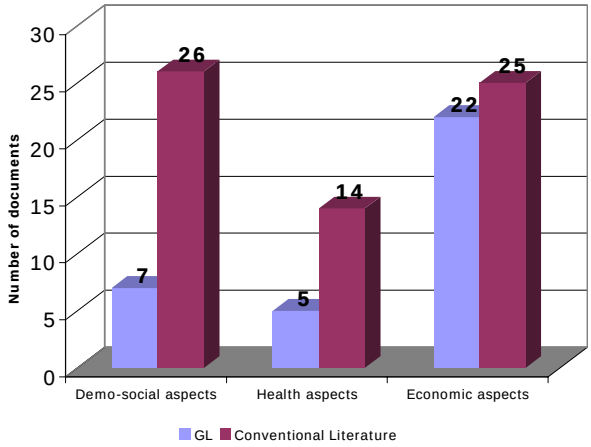


Fig. 5. – Distribution of the cited documents by topic and type of publication

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		**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**	
	Frequency No.	**Cumulative frequency %**	**Frequency No.**	**Cumulative frequency %**	**Frequency No.**	**Cumulative 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 available	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

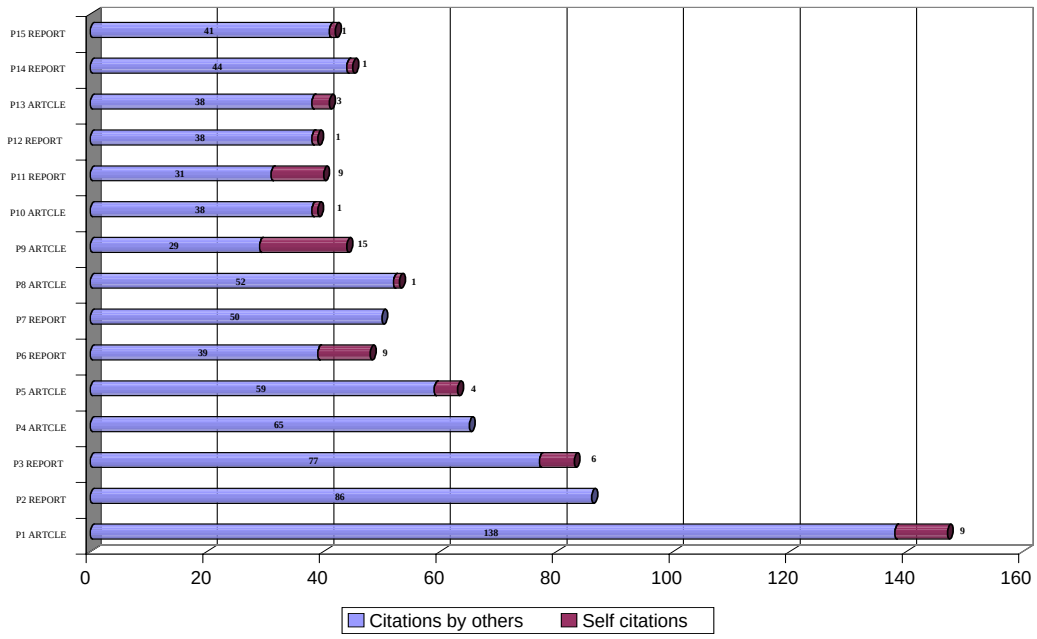
			Grey literature		Conventional literature	
Documents	**Type of document cited**	**Total number of citations**	**No.**	**%**	**No.**	**%**
P1	ARTICLE	147	81	55.1	66	44.9
P2	REPORT	86	54	62.8	32	37.2
P3	REPORT	83	61	73.5	22	26.5
P4	ARTICLE	65	14	21.5	51	78.5
P5	ARTICLE	63	18	28.6	45	71.4
P6	REPORT	48	40	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			
Documents	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.

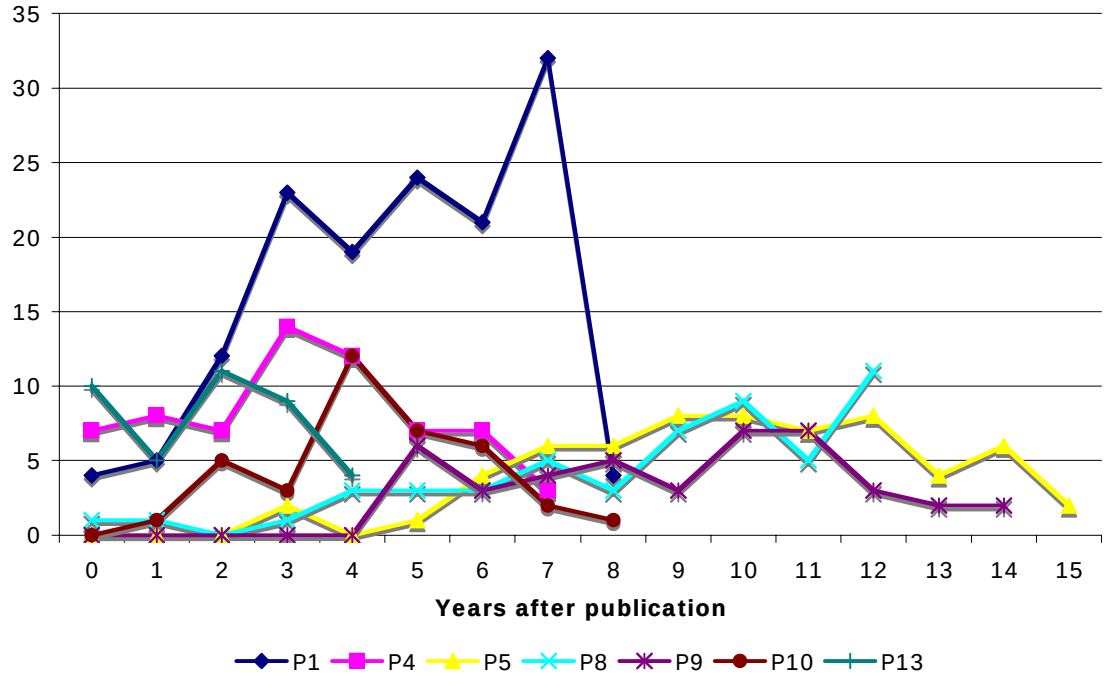


Fig. 7. - Distribution of citations over time in conventional documents

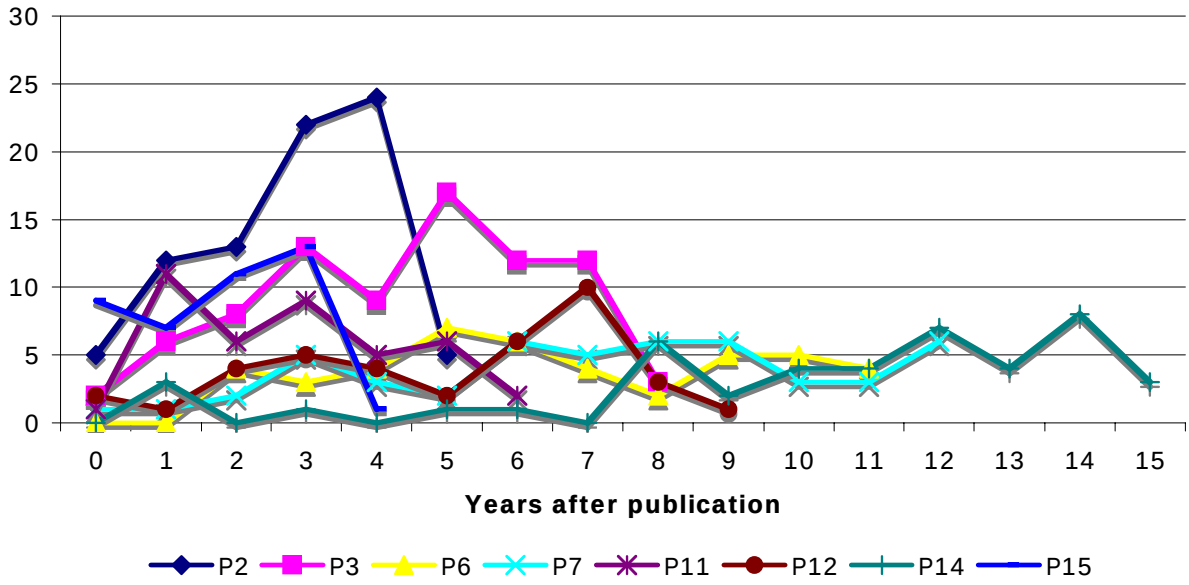


Fig. 8. - Distribution of citations over time in GL documents

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 o citations	Total number o 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 an 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 an 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 OECI area / A. Borsch-Supan. Institut für Volkswirtschaft und Statistik, Discussion pape series No. 512, 1995.	48	9
P7-R	The consequences of population aging on private pension fund saving and asse 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 / Maxim Fougère and Marcel Mérette. Economic Studies and Policy Analysis, Division Department of Finance, Workin 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

https://doi.org/10.26069/greynet-2026-000.604-tgj

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

Primož Južnič
(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.

* First published in the GL9 Conference Proceedings, February 2008

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.

https://doi.org/10.26069/greynet-2026-000.604-tgj

Digital Documents in Grey Literature: New Challenges *

Aleksandr V. Starovoitov, Lev G. Titarev, Yuri M. Bogdanov, and Leonid P. Pavlov
(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.

* First published in the GL9 Conference Proceedings, February 2008

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.

References

- VNTIC, Scientific and Technical Information Centre of Russia, http://www.vntic.org.ru/
- GOST 7.32 – 2001, Scientific report standard (System of standards on information, librarianship and publishing) http://www.bmstu.ru/~rl1/courses/electr/KursRabotaRK6_3k/KURS%20RAB%20RL6%204K%20KD/Metoda%20PDF%20KD/GOST%207_32_91/GOST-7.32-91.pdf

https://doi.org/10.26069/greynet-2026-000.604-tgj

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 750 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!

For more information visit our website: www.greyliterature.org
or contact us at: greylithelp@nyam.org

The New York Academy of Medicine

At the heart of urban health since 1847

https://doi.org/10.26069/greynet-2026-000.604-tgj

Quarterly News Report on Grey Literature

Sample Copy

ISSN 1876-2476

GL Compendium

GREY LITERATURE NETWORK SERVICE

Launch GL Compendium

GL Compendium is a quarterly news report on grey literature. It is designed to guide practitioners and inform researchers and educators in their search for and access to grey literature.

GLC is an international, multi-disciplinary one-stop source on valuable grey literature. It taps into web-based resources and collections, harvests quality information and makes it available on your desktop in digital format.

GLC is a state of the art serial publication compiled and edited by GreyNet, a leader in the field of grey literature and networked information for over 15 years. Traditionally, GreyNet focused on research and education in grey literature.

GLC

Now, the GL Compendium expands and enhances GreyNet's service to practitioners - librarians, subject specialists, policy and decision makers, information managers, and others. GL Compendium provides a unique current awareness resource for LIS professionals.

GLC grew out of GreySource, a popular web-based product. However, while GLC maintains the same classification scheme, it moves beyond simple hyperlinking by tunnelling deeper into the document and content level of grey literature.

Contents

GreySource Frontrunner 2008

Information for Practice (IP) edited by Professor Gary Holden, New York University; Silver School of Social Work has been voted GreySource Frontrunner 2008. IP's mission is to help social service professionals throughout the world conveniently maintain an awareness of news regarding the profession and emerging scholarship. Prof. Holden's work on "Information for Practice" is truly an example of best practice in the field of grey literature. http://www.nyu.edu/socialwork/ip/

Editorial Address

GreyNet
Grey Literature Network Service
Javastraat 194-HS
1095 CP Amsterdam
The Netherlands

Tel/Fax 31-20-331.2420
info@greynet.org
www.greynet.org/glcompendium.html

GLC is for keepsake!

GL Compendium enables you to:

- ✓ Keep an eye out for new sources in grey literature
- ✓ Keep informed on organizations in the field
- ✓ Keep up with grey literature in your subject area
- ✓ Keep ahead on policy issues affecting grey literature
- ✓ Keep abreast of collections and resources in your discipline
- ✓ Keep in touch with repositories housing grey literature
- ✓ Keep tabs on uses and applications of grey literature, and
- ✓ Keep down costs, taking advantage of a low subscription rate

https://doi.org/10.26069/greynet-2026-000.604-tgj

Colophon

CIP
GL Compendium (GLC): A Quarterly News Report on Grey Literature / GreyNet, Grey Literature Network Service. – Amsterdam : TextRelease, Volume 1, Autumn Issue *(Forthcoming)*, 2008.

The British Library, INIST-CNRS, and the New York Academy of Medicine are GLC corporate authors and GreyNet Associate Members. This serial publication appears four times a year - in winter, spring, summer, and autumn. The contents of each issue follow in sequence. The subject areas are based on a derivative of the COSATI (American) and SIGLE (European) Classification Systems. And, the LIS sections conform to GreyNet's involvement in the field of grey literature. - ISSN 1876-2476 (PDF)

Editor:
Dr. Dominic J. Farace
GreyNet, Grey Literature Network Service

Technical Editor:
Jerry Frantzen, TextRelease

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

Subscription Fee: €160
All in price covers 4 consecutive issues

Publisher Address:
(For back issues, advertising, and subscriptions)

TextRelease
Javastraat 194-HS
1095 CP Amsterdam
The Netherlands

T/F +31 (0) 20 331.2420
info@textrelease.com
http://www.textrelease.com/glpublications.html

GLC Subject Classification:

00 - General, Multidisciplinary
01 - Aeronautics
02 - Agriculture, Forestry, Fisheries, Veterinary Sciences
03 - Environmental Pollution, Protection and Control
04 - Humanities (History, Philosophy, Religion, Etc.)
05 - Social Sciences (Economics, Information, Psychology, Etc.)
06 - Biological and Medical Sciences
07 - Chemistry
08 - Earth and Atmospheric Sciences
09 - Electronics, Electrical Engineering, Computer Science
10 - Energy & Power
11 - Materials
12 - Mathematical Sciences
13 - Mechanical, Industrial, Civil & Marine Engineering
14 - Methods & Equipment
15 - Military Sciences
16 - Missile Technology
17 - Navigation, Communication, Detection, Etc.
18 - Science and Technology
19 - Ordnance
20 - Physics
21 - Propulsion & Fuels
22 - Space Technology

https://doi.org/10.26069/greynet-2026-000.604-tgj

GL 10

Designing the Grey Grid for Information Society

Tenth International Conference on Grey Literature

Science Park Amsterdam, Netherlands

December 8 - 9, 2008

Conference Program

DAY ONE

9:00-10:30

OPENING SESSION

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

Opening Address, *Arie Jongejan, Chief Executive Officer Swets, The Netherlands*

Keynote Address, *Brian Hitson, Associate Director OSTI-DOE, United States*

Rejoinder, *Elizabeth Newbold and Jennie Grimshaw, The British Library, United Kingdom*

11:00-12:30

SESSION ONE – INSTITUTIONAL REPOSITORIES AND GREY LITERATURE

Chair, *Anne Asserson, University of Bergen, Norway*

Grey Literature in the Czech Republic
Petra Pejsova, Martina Pfeiferova, State Technical Library, Czech Republic

Towards an Institutional Repository of the Italian National Research Council: A Survey on Open Access Experiences
Daniela Luzi, Rosa Di Cesare, Roberta Ruggieri and Loredana Cerbara, Institute of Research on Population and Social Policies, IRPPS-CNR, Italy

Grey literature in French Digital Repositories: A Survey
Joachim Schöpfel and Christiane Stock, INIST-CNRS, France

INIS Grey Literature - A Trusted Nuclear Full Text Repository
Seyda Rieder, International Atomic Energy Agency, IAEA, Austria

13:30-15:00

SESSION TWO – GREY LITERATURE IN BIOMEDICAL COMMUNITIES

Chair, *Elizabeth Newbold, The British Library, United Kingdom*

Recommendations for finding the gold: Optimizing efforts to identify the Grey Literature on Public Health Effectiveness
Maureen Dobbins and Paula Robeson, Health-evidence.ca; McMaster University, Canada

Knowledge and Use of Grey Medical Literature by Medical and Paramedical Staff at University/Research Associated Medical Institutions in Developing Countries
Emmanuel U. Onyekwelu, Joan Carlos, Rodriguez David, and Margareta Gonzalez Royal Victoria Teaching Hospital, The Gambia

Harnessing the Power of the Veterinary Grey Literature
Joe Jaros and Esther Carrigan, Texas A&M University [et al.], United States

Grey Literature and Development: The Non-Governmental Organization in Action
Lynne Marie Rudasill, University of Illinois at Urbana-Champaign, United States

15:30-17:00

POSTER PRESENTATIONS

Chair, *Dr. Primož Južnic, University of Ljubljana, Slovenia*

(See Listing of Posters, below)

GL10 Program and Conference Bureau
TextRelease
Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com ▪ **conference@textrelease.com**
Tel/Fax +31 (0) 20-331.2420

https://doi.org/10.26069/greynet-2026-000.604-tgj

GL 10

Designing the Grey Grid for Information Society

Tenth International Conference on Grey Literature

Science Park Amsterdam, Netherlands
December 8 - 9, 2008

Conference Program

DAY TWO

09:00-10:30

POSTER PRESENTATIONS

Chair, *Dr. Primož Južnic, University of Ljubljana, Slovenia*

▪ **ISPESL** (Italy) ▪ **SNDT** (India) ▪ **ISFOL** (Italy) ▪ **UNISWA** (Swaziland) ▪ **UFSC** (Brazil) ▪ **OPI** (Poland) ▪ **KISTI** (Korea) ▪
▪ **Salamanca University** (Spain) ▪ **VNTIC** (Russia) ▪ **University of Calgary** (Canada) ▪ **University of Malaya** (Malaysia) ▪ **Etc**

Note *The order of the Posters, Title information, and format for their presentation is forthcoming.*

11:00-12:30

SESSION THREE – LEGAL ASPECTS, INTELLIGENCE, AND TEXT MINING IN GREY LITERATURE

Chair, *Christiane Stock, INIST-CNRS (France)*

Green Light for Grey Literature? Orphan Works, Web-Archiving and other Digitization Initiatives – Recent Developments in U.S. Copyright Law and Policy
Tomas A. Lipinski, School of Information Studies; University of Wisconsin, United States

Legal deposit practices regarding grey audiovisual and multimedia materials carrying Creative Commons licenses and those carrying traditional copyright at the Library of Congress
Debbie L. Rabina, Pratt Institute; School of Information and Library Science, United States

The "Grey" Intersection of Open Source information and Intelligence
June Crowe and Thomas S. Davidson, Open Source Research Group; IIA, Inc., United States

Grey Literature for Natural Language Processing: A Terminological and Statistical Approach
Laura Cignoni, Gabriella Pardelli, and Manuela Sassi, *Istituto di Linguistica Computazionale, CNR, Italy*

13:30-15:00

SESSION FOUR – GREY LITERATURE IN RESEARCH

Chair, *Daniela Luzi, CNR-IRPPS (Italy)*

Do High-Energy Physics Scholars read Scientific Journals?
Anne Gentil-Beccot, European Organization for Nuclear Research, CERN, Switzerland

Grey Literature produced and made available by Universities – Helping future Scholars or Copycats?
Primož Južnic, University of Ljubljana, Department of Library and Information Science and Book Studies, Slovenia

Interest - INTERoperation for Exploitation, Science and Technology
Keith G Jeffery, Science and Technology Facilities; Council Rutherford Appleton Laboratory, United Kingdom
Anne Asserson, University of Bergen, Research Department, Norway

OpenSIGLE, Home to the GreyNet Community and its GL Collection: First Results and a Research Proposal
Dominic Farace and Jerry Frantzen; Grey Literature Network Service, Netherlands
Joachim Schopfel, Christiane Stock, and Nathalie Henrot; INIST-CNRS, France

15:15-16:00

CLOSING SESSION - REPORTS FROM CHAIRPERSONS, CONFERENCE EVALUATION, AND FAREWELL

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

GL10 Program and Conference Bureau
TextRelease
Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com ▪ conference@textrelease.com
Tel/Fax +31 (0) 20-331.2420

About the Authors

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

Elly Dijk graduated from 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 working group Repository Managers of the SURFshare programme (2007-2010). Elly Dijk is also a 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

Ulrich Herb studied Sociology at Saarland University in Germany. He is employed at Saarland University and State Library, which is not only the local University and State Library for Saarland but also the special subject collection library for the discipline Psychology in Germany. He headed different projects dealing with free of charge electronic scientific information in the field of Psychology like the virtual psychology library and Digital Psychology Information DPI. In addition to the above mentioned projects he engaged in a number of initiatives dealing with the aspects of nationwide information supply both disciplinary and interdisciplinary. Today his major tasks are digital publishing in general, development and management of electronic publishing services (regional at Saarland University and supra-regional for the psychologist's community in Germany), alternative publishing models and Open Access.
Email: u.herb@sulb.uni-saarland.de

Primož Južnič is an associate professor at the Department of Library and Information Science and Book Studies at Faculty of Arts, University of Ljubljana (Slovenia). His main area of research and interest is bibliometrics, collection management and LIS education. He teaches the following courses: Bibliometrics, Special libraries, and Collection Management. Before starting his university career, he was a heading different special and academic libraries and information/computer centres. He was also working at the European Commission, for three years, as the seconded informatics expert.
Email: primoz.juznic@ff.uni-lj.si

Matthias Müller studied biology at Saarland University. He is deputy director and member of the electronic publishing group of the Saarland University and State Library.

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

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

https://doi.org/10.26069/greynet-2026-000.604-tgj

Notes for Contributors

Non-Exclusive Rights Agreement

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

Submission Requirements

Manuscripts should not exceed 15 double-spaced typed pages. The size of the page can be either A-4 or 8½x11 inches. Allow 4cm or 1½ inch from the top of each page. Provide the title, author(s) and affiliation(s) followed by your abstract, suggested keywords, and a brief biographical note.
A printout or PDF of the full text of your manuscript should be forwarded to the office of TextRelease. A corresponding MS Word file should either accompany the printed copy or be sent as an attachment by email. Both text and graphics are required in black and white.

REFERENCE GUIDELINE

General

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

Specific

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

Examples:

Youngen, G.W. (1998), Citation patterns to traditional and electronic preprints in the published literature. - In: College & Research Libraries, 59 (5) Sep 1998, pp. 448-456. - ISSN 0010-0870

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

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

Review Process

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

Journal Publication and Article Deposit

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

https://doi.org/10.26069/greynet-2026-000.604-tgj
TGJ Volume 4, Number 2 2008

An International Journal on Grey Literature

'ACCESS TO GREY IN A WEB ENVIRONMENT'

SUMMER 2008 – TGJ VOLUME 4, NUMBER 2