Sixteenth International Conference on Grey Literature

The Library of Congress, Washington D.C. USA, 8-9 December 2014

GREY LITERATURE LOBBY: ENGINES AND REQUESTERS FOR CHANGE

Conference Proceedings

ISSN 1386-2316

Host and Sponsors

GL16 Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com · conference@textrelease.com
Tel/Fax +31-20-331.2420

CIP

GL16 Conference Proceedings

Sixteenth International Conference on Grey Literature : Grey Literature Lobby: Engines and Requesters for Change, Library of Congress, Washington D.C., 8-9 December 2014 / compiled by D. Farace and J. Frantzen ; GreyNet International, Grey Literature Network Service. Amsterdam : TextRelease, March 2015. – 164 p. – Author Index. – (GL-conference series, ISSN 1386-2316 ; No. 16).

FEDLINK-Library of Congress (USA), CVTISR (SK), DANS (NL), EBSCO (USA), Inist-CNRS (FR), ISTI-CNR (IT), KISTI (KR), NIS-IAEA (AT), NTK (CZ), and NYAM (USA) are Corporate Authors and Associate Members of GreyNet International. This conference proceedings contain the full text of 17 papers presented during the two days of plenary and poster sessions. The papers appear in the same order as in the conference program book. Included is an author index with the names of the contributing authors/coauthors along with their biographical notes. A list of more than 50 participating organizations as well as sponsored advertisements are likewise published in these conference proceedings.

ISBN 978-90-77484-23-4

Foreword

GREY LITERATURE LOBBY

Engines and Requesters for Change

Decision and policy makers need to be informed on the value and wealth of grey literature, thus legitimizing further investments in this field of information. Lobbying grey literature has its very roots in this international conference series, which has grown and rallies over the past two decades by promoting research and publishing their results. The grey literature lobby seeks to guarantee that the interests of a diverse and widespread community of information professionals and practitioners are served.

Like other fields in library and information science, technology is one of the primary engines driving change in grey literature. However, there are other engines for change that are needed to further sustain and develop this field of information. Policy development and economic stimulus are two such challenges now facing the community.

Changes in the production and sharing of knowledge, changes in the requirements for storage, access, and preservation of grey literature, as well as ensuing change in the demands of users require a concerted effort and response on the part of the grey literature community. New stakeholders and net-users must be summoned and heard, because their requests will further fuel the right kind of debate needed to open the wealth of grey literature to wider audiences.

The social impact of grey literature will be judged by the value placed on its public service. Some seventeen papers from authors and researchers from fifteen countries worldwide are harvested in these Proceedings.

Dominic Farace ***Amsterdam,***
GREYNET INTERNATIONAL MARCH 2015

GL16 Conference Sponsors

FEDLINK, USA
Federal Library Information Network
Library of Congress

CVTISR, Slovak Republic
Slovak Centre of Scientific and Technical Information

KISTI, Korea
Korea Institute of Science and Technology Information

Grey Literature Strategies, Australia
Swinburne University of Technology

EBSCO Publishing, USA

NIS-IAEA, Austria
Nuclear Information Section
International Atomic Energy Agency

GL16 Conference Sponsors

ISTI-CNR, Italy
Institute of Information Science and Technologies
National Research Council, CNR

NTK, Czech Republic
National Library of Technology

NYAM, USA
New York Academy of Medicine

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

IIa, USA
Information International Associates, Inc.

DANS, The Netherlands
Data Archiving and Networked Services,
Royal Netherlands Academy of Arts and Sciences

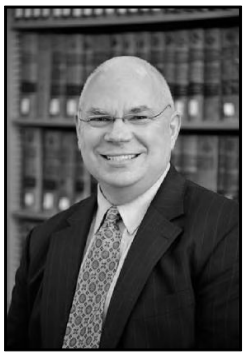
GL16 Conference Moderators and Chair

Moderator Day One

Amanda J. Wilson, Director
National Transportation Library
U.S. Department of Transportation

Amanda became Director of the National Transportation Library (NTL), an all-digital library, in December 2006. In her time at NTL, Ms. Wilson has focused on building collections, increasing awareness of NTL collections and services, and building a national network of transportation libraries. Under her leadership, NTL has engaged in new initiatives in pursuit of these goals, establishing the Transportation Librarians Roundtable (a monthly webinar series for transportation librarians), providing national coordination of projects between and among regional library networks in transportation, engaging in federal science and technology information leadership through CENDI, and expanding NTL's data programs, including development of a USDOT data catalog and National Transportation Data Archive, and active involvement in the Research Data Alliance and the Federal Statistical Community of Practice and Engagement.

Amanda.Wilson@dot.gov



Conference Chairman

Blane K. Dessy,
Deputy Associate Librarian
Library of Congress

Blane is a Deputy Associate Librarian with a focus on Planning and Project Management in Library Services at the Library of Congress. He was the former Executive Director of the Federal Library Information Network at the Library of Congress and was instrumental in establishing the FedGrey Working Group. In 2010, Blane delivered the Keynote Address and at the Twelfth International Conference on Grey Literature. Prior to his appointment at the Library of Congress, Blane had been Director of Libraries at the United States Department of Justice and the first Executive Director of the National Library of Education. Blane came to the Federal Government after working as a State Librarian, Deputy State Librarian, library consultant, and public library director. He was also an adjunct instructor in Management at the Catholic University, School of Library and Information Science.

bdes@loc.gov

Moderator Day Two

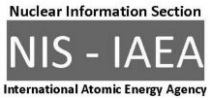
Michele Masias,
Chief Librarian, Justice Libraries
U.S. Department of Justice

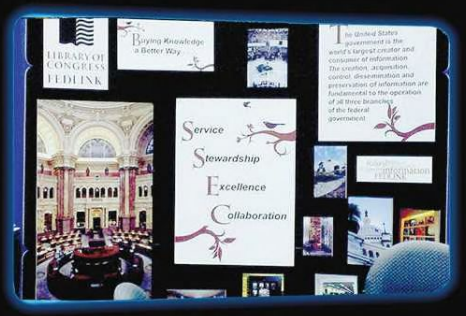
Michele serves as the Chief Librarian for the Civil and Criminal Divisions, managing four Justice Libraries. She has over 20 years of library experience in positions that have included serving as the Law Librarian at the Executive Office of the President, the Library Manager at the Defense Technical Information Center, and as a Technical Information Specialist at the Department of Interior Bureau of Reclamation Library. She is currently an elected member of the Fedlink Advisory Board and a Lecturer at the Catholic University of America. She has authored several articles for library publications such as Colorado Libraries, Law Library Lights, Information Outlook, and the U.S. Attorneys' Bulletin. Michele received her MLS from Emporia State University. She's a member of the American Association of Law Libraries and the Law Librarians Society of Washington, D.C.

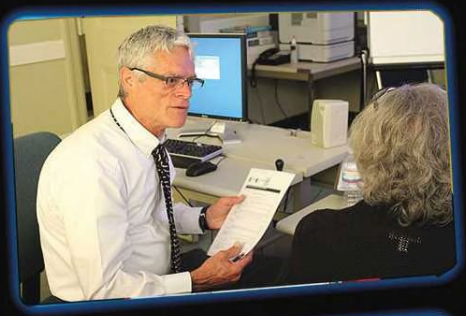
Michele.Masias@usdoj.gov

Table of Contents

GL16 Program Committee

	Blane Dessy **Chair** Library of Congress	United States
	Danielle Aloia New York Academy of Medicine, NYAM	United States
	Ján Turňa Slovak Centre of Scientific and Technical Information, CVTI SR	Slovak Republic
	Stefania Biagioni Institute of Information Science and Technologies, ISTI-CNR	Italy
	Kiseok Choi Korea Institute of Science and Technology Information, KISTI	Korea
	Petra Pejšová National Library of Technology, NTK	Czech Republic
	Dobrica Savic Nuclear Information Section, NIS-IAEA	Austria
	Joachim Schöpfel University of Lille 3	France
	Christiane Stock Institut de l'Information Scientifique et Technique, INIST-CNRS	France
	Debbie Rabina Pratt Institute, School of Information and Library Science	United States
	Tomas A. Lipinski Kent State University	United States
	Dominic Farace Grey Literature Network Service, GreyNet International	Netherlands





FEDLINK

an organization
of federal agencies
working together
to achieve optimum use of
resources and facilities
of federal libraries
and information centers
by promoting
common services,
coordinating and sharing
available resources, and
providing continuing
professional education.

LIBRARY OF CONGRESS
FEDLINK

101 Independence Ave, SE ~ Washington, DC 20540-4935
FEDLINK Main Number (202) 707-4800
FEDLINK Hotline (202) 707-4900

The Intersect of Cultural Studies with other focus areas in the HDIAC Database

June Crowe and Crystal Sherline

Information International Associates, Inc. United States

Our bibliometric research will examine a select set of documents in the Homeland Defense and Security Information Analysis Center's (HDIAC) collection. The focus will be on documents in the "Cultural Studies" focus area. There are over 210,000 documents in the HDIAC collection, and much of it is grey literature. Staff use a template that includes bibliographic information, keywords, task areas, and other descriptive information to catalog items for inclusion in HDIAC's database. In staff discussions, it was discovered that some focus areas intersect with other focus areas. For example, Cultural Studies overlaps with the Medical, Alternative Energy, Critical Infrastructure Protection, and Homeland Defense & Security focus areas. The purpose of this research is to examine the intersects that Cultural Studies has with the other HDIAC focus areas by examining the keywords and task area terms that include the term "Cultural Studies." This process will assist in clarifying the subject key words used to identify cultural studies and facilitate the tagging, acquisition, and discovery processes. It will also give staff a model they can use to gain a deeper understanding of how culture intersects with other disciplines. A bibliometric study will be conducted to evaluate the HDIAC database, Scientific and Technical Analysis and Research Tool (START), and quantify the Activity Index (AI) of the HDIAC database.

$$AI_1 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Publications\ in\ START}$$

and

$$AI_2 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Cross\ \ Publications\ in\ START}$$

This bibliometric research will allow HDIAC staff to quantify the multidisciplinary nature of Cultural Studies and provide a vantage point for a qualitative study that examines the intersect with other focus areas in HDIAC.

Keywords: Scientific and technical data, Gap Analysis, Grey Literature, Data Discovery, Taxonomy, Bibliometrics

Introduction

The Homeland Defense and Security Information Analysis Center (HDIAC) is located in Oak Ridge, Tennessee and is a knowledge center of excellence for the Department of Defense (DoD) Homeland Defense and Security (HLDS) scientific and technical information (STI). Previously known as the Chemical, Biological, Radiological and Nuclear Information Analysis Center (CBRNIAC), the HDIAC has eight focus areas: Homeland Defense & Security, Critical Infrastructure Protection (CIP), Weapons of Mass Destruction (WMD), CBRN Defense, Biometrics, Medical, Cultural Studies, and Alternative Energy. The last four topic areas are new to HDIAC. Information resources in these areas are collected, analyzed, synthesized, and disseminated to the DoD community. Additionally, the HDIAC produces technical reports, journal articles, and web content, as required in these areas. HDIAC's goals are to: use the best knowledge and technical expertise from the government, industry, and academia to solve difficult scientific and technical problems in HLDS; serve as a ready tool for strategic, operation, and tactical organizations in DoD and the broader homeland defense community to resolve issues; build a community of homeland defense SMEs and provide long-term STI corporate memory for DoD; and build networks to reduce duplicate information holdings in the research and engineering communities (STI –reuse).

Research Method

To examine the HDIAC collection and aid in the development of a richer literature taxonomy, the researchers examined Task Keywords, i.e., the category of the HDIAC scope areas (Alternative Energy, Cultural Studies, WMD, etc.). This assessment allows for quantification of the juxtaposition of the Cultural Studies literature with literature in the other scope areas in the collection, giving an Activity Index (AI).

$$AI_1 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Publications\ in\ START}$$

In order to review the literature, an evaluation of the HDIAC START database was performed, and metrics were examined. First, the START database metrics were retrieved by searching Task Keywords for the eight HDIAC scope areas. Once metrics were retrieved for the individual scope areas, metrics were collected for areas where Cultural Studies and other scope areas specifically intersected. In order to obtain these metrics, a search in START was formulated by searching Task Keywords "Cultural Studies" and "Alternative Energy." This was also done for all the other six scope areas, giving an AI_2.

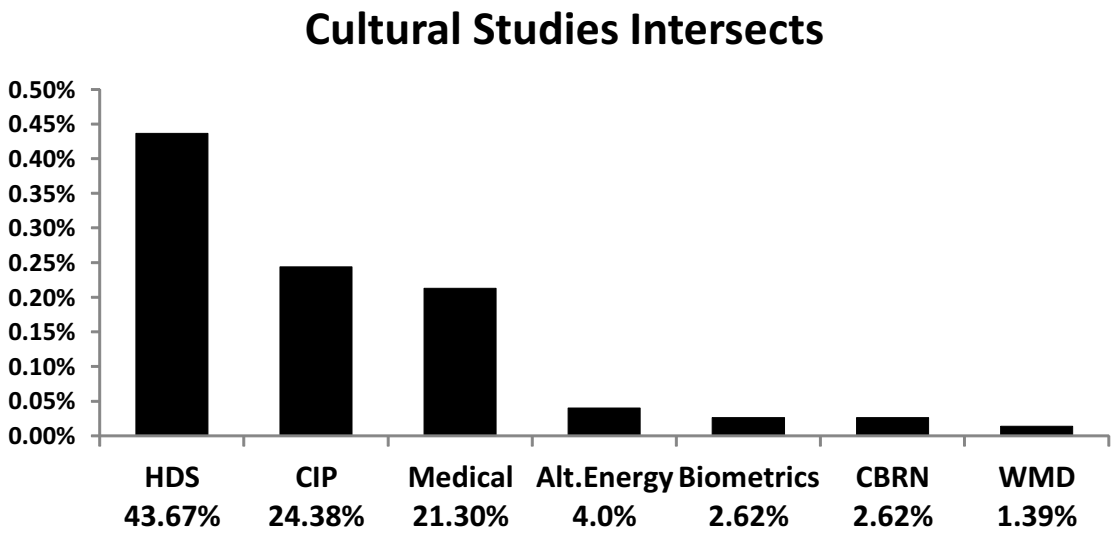
$$AI_2 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Cross\ \ Publications\ in\ START}$$

Bibliometric Research Analysis

The database that HDIAC inherited from the previous contractor contained very few records, if any, on the four new focus areas (Cultural Studies, Alternative Energy, Biometrics, and Medical). The original database focused on chemical, biological, radiological, and nuclear subject areas. In essence, there were almost no documents on the new focus areas. HDIAC staff developed a new taxonomy from the performance work statement that provided an overarching framework for collecting information on the various subject domains at a first and second level. The current need is for a more in-depth taxonomy for each focus area. Additionally, our research will assist with gap analysis in collection development in the eight focus areas. Currently, the collection consists of grey literature, journal articles, technical reports, maps, books, and data on various topics. Since November 11, 2013, HDIAC has added 8,880 records related to the focus areas.

Cultural Studies documents comprise 14.82% of the additions or 1,316 items. Cultural Studies is a cross disciplinary domain that intersects with Alternative Energy, Critical Infrastructure Protection (CIP), Homeland Defense and Security (HLDS), Biometrics, Medical, and Weapons of Mass Destruction (WMD). For example, "megacities" is a category under Cultural Studies that also intersects with CIP, HLDS, Alternative Energy, Medical, and Biometrics. Depending on the document, this topic could also easily intersect with CBRN and WMD. Figure 1 shows how Cultural Studies intersects with the other domains. As can be seen in Figure 1, HLDS, CIP, and Medical are the top three focus areas that have intersects with Cultural Studies. The total percentage of Cultural Studies documents with cross over task areas at this point in time is 49.24%, almost half of the content that has been cataloged.

Figure 1: Cultural Studies Intersects with other Focus Areas



In addition to task area terms, the authors plan to examine the subject keywords that describe the eight focus areas in an attempt to standardize and develop a more detailed subject thesaurus for the HDIAC collection. The HDIAC START system's template for cataloging includes bibliographic information, a distribution statement, copyright information, abstract, document type, corporate authorship, task keywords, and subject keywords. The task keywords are in effect the subject domain terms or HDIAC scope area. The subject keywords that have been used have been taken directly out of the statement of work pertaining to this contract. Subject key wording has become more important in search efforts, as the collection has grown in both breadth and depth. The subject keywords provided by the contract no longer encapsulate the documents being collected. Therefore, is has become imperative that a more accurate, homogenous method to categorize the information is created.

Our research will further develop the existing taxonomy for each of the domains to facilitate catalog tagging and searching. For example, under the heading "Behavior," there is a subcategory of "human adaptation and response to perturbations," which will have further subcategories that may include climate change, adaptation to food stress, or civil unrest.

Multidisciplinary Example

Figure 2 is a document cataloged in HDIAC's START database. This document was chosen as an example because it does not fit into just one of the HDIAC's scope areas. The document is cataloged as having the following task keywords: "Cultural Studies," "Alternative Energy," and "Critical Infrastructure Protection."

Figure 2. Multidisciplinary Example Document

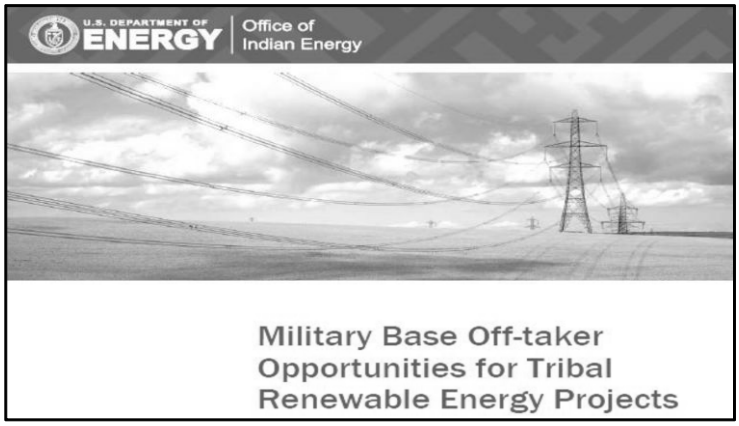


Figure 3 is the document's cataloging output in the START database. The document is a multidisciplinary document and has been cataloged with three different task key words (HDIAC scope areas), yet the subject keywords are vague and only contain secondary subject headings related to some of the task keywords. Ideally, the subject headings for this document would include: diplomacy, communication, land ownership, socio-economics, and environmental resources.

Figure 3. HDIAC Catalog Record

IAC Number: HD-224134
AD Number:
Date Received: 2014-03-31
Title: Military Base Off-taker Opportunities for Tribal Renewable Energy Projects.
Site Holding: HD
SiteHolding Number:
Media Type: PDF
CB Collection: UA
Author(s): Nangle, J.
Corp Source Code: 424651
Corp Author Name: NATIONAL RENEWABLE ENERGY LAB GOLDEN CO
Descriptive Note: White Paper
Publish Date: 2013-05-01
Page Count: 19
Country: US
Contract Number:
Project Number:
Report Number: DOE/IE-0018
Distribution Code: PUBLIC
Distribution Statement: Approved for Public Release; Distribution Unlimited.
Abstract: Tribal lands represent approximately 5% of the national renewable energy technical potential (Doris, Lopez, and Beckley 2013), and their close proximity to many DOD installations makes the Tribes potential partners in helping the DOD meet its renewable energy goals. The benefits to the Tribes include additional income streams and workforce development in green jobs. This white paper surveys DOD installations that could have an interest in the purchase of energy from renewable energy projects on tribal lands. Tribes, in the right context, could supply local renewable energy to meet military demands, which would provide additional sources of energy to the installations, helping them meet federal requirements for renewable energy purchases and greenhouse gas reduction. Tribal members could benefit from additional income from land leases and energy sales, as well as workforce development and maintaining the renewable energy projects. Identification of likely purchasers of renewable energy is a first step in the energy project development process, and this paper aims to identify likely DOD electricity customers that tribal commercial-scale projects could serve.
Supplemental Notes:
Subject Keywords: BIOMASS; DEPARTMENT OF ENERGY (DOE); GEOTHERMAL; HYDROPOWER; INSTALLATION ENERGY POTENTIAL; PHOTOVOLTAIC (PV); PV-RURAL; PV-URBAN; TRIBAL LAND
Task Keywords: ALTERNATIVE ENERGY CRITICAL INFRASTRUCTURE PROTECTION CULTURAL STUDIES OTHER GOVERNMENT AGENCY TEMS-WS 2014-07-29

Taxonomy Development

Since Cultural Studies is one of the four new focus areas, our basic overarching taxonomy was devised using the performance work statement. It was quickly recognized that a more detailed and standardized taxonomy was needed for both cataloging and search purposes. Some preliminary research into the use of other existing taxonomies was conducted. "For example, the Defense Technical Information Center (DTIC) has a technology taxonomy while the Standard Subject Identification Code (SSIC) is the standard for all DoD information, including memorandums and records management. Similarly, the Library of Congress Classification (LOCC) is a commonly used general purpose system. However, taxonomies inevitably have a central theme that guides how the tree structure is arranged."[1]

Our first problem was to determine what subject key terms had already been used in conjunction with the Cultural Studies task tag. An SQL query was performed to find those terms and rank them by frequency of use. Since Cultural Studies is multidisciplinary, this query can provide crossover terms that may have been used from any other task areas. Secondly, DTIC's taxonomy,[2] which is available online at http://dtic.mil/dtic/stresources/techreports/dticSearchTools/thesaurus_desc.html, needed to be evaluated. DTIC's taxonomy did not provide the level of detail needed for the scope of Cultural Studies. The Library of Congress' "Subject Heading List"[3] for Culture, Anthropology, Social Science, and Human Geography, and "Military Culture Factors" was also consulted. We are in the process of evaluating all of the relevant cultural terms from these two sources. Other taxonomies will also be researched. After relevant terms from all these sources are selected, we will look into possible programs

that could integrate those terms with the broad overarching terms. Figure 4 is an example of what the tree may look like. The program will be run periodically to update the taxonomy. After the Cultural Studies taxonomy is completed, the same gap analysis and integration of new terms will be conducted for the other focus areas.

Figure 4. Example Taxonomy Tree

Dominant Values & Beliefs (Broad Term)

- Social Values & Beliefs (second level of Broad Term)
- Religion (broad term from Military taxonomy)
  - Major tenets (more specific terms from military taxonomy)

Conclusion

After reviewing and quantifying the AI of the HDIAC's database, START, several deductions can be made. Firstly, the HDIAC scope area of Cultural Studies is a multidisciplinary area. The AI index shows that nearly half of the Cultural Studies documents cataloged in the database intersect with at least one other scope area. Because of the multidisciplinary status of the Cultural Studies literature being collected, it has become increasingly more important to develop a living, growing taxonomy specific to the HDIAC collection. Existing taxonomies within the Library of Congress and the DoD, although useful as a stepping stone, are not specific enough to delineate the HDIAC collection. Secondly, with projected collection growth of HDIAC literature at 5% per year, it is imperative a working framework for scope area taxonomies, which will create standards not just for collection but for search and retrieval, be created. The evolving taxonomies also benefit the end user by ensuring a more accurate return. Lastly, developing an organic taxonomy for HDIAC's scope areas allows literature gaps to be recognized and addressed, which in turn ensures a more robust collection.

Endnotes

[1] Dr. Geoffrey P Malafsky and Newman, Brain. Organizing Knowledge with Ontologies and Taxonomies, TECHi2 , Fairfax, VA. http://techi2.com/download/Malafsky%20KM%20taxonomy_ontology.pdf

[2] DTIC Online. http://dtic.mil/dtic/stresources/techreports/dticSearchTools/thesaurus_desc.html

[3] Library of Congress Subject Headings. The Library of Congress. 18 October 2014: http://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html

IIa
Information
International
Associates

Navigators in a Sea of Data

For 23 years, IIa has provided quality information services to advance our clients' mission. By applying information science skills in combination with information technology, we help organizations handle massive amounts of information and make it useable.

IIa makes information an asset for your organization.

Corporate Office:
104 Union Valley Road
Oak Ridge, TN 37830
Phone: 865- 481-0388

D.C. Area Office:
417 West Broad Street
Suite 201
Falls Church, VA 22046
Phone: (703) 407-5770

www.iiaweb.com

Mapping Italian Grey Communities: What Is There Beyond The Academy?

Silvia Giannini and Stefania Biagioni,
CNR, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Italy

Sara Goggi and Gabriella Pardelli,
CNR, Istituto di Linguistica Computazionale, "Antonio Zampolli", Italy

1. Introduction

The following title was published on an influential Italian newspaper, *La Stampa*, on November 7, 2013: "Tra i tesori della 'Letteratura Grigia' un'Eneide in napoletano del '600". The article is about the presentation of the "Fondo De Mauro" on the Italian Network of Popular Culture [1]: this fund originates from a private collection built up in several decades and donated to the Network by Tullio De Mauro[1] and his wife Silvana Ferreri in 2011; it is made up of thousands of books, brochures, pamphlets relating to Italian dialects and minority languages.
In the field of healthcare, while browsing on the web, we found the title of a PhD thesis of last year: "Letteratura Grigia nelle meta-analisi delle prove ripartite con scelta casuale degli interventi di sanità".
In the legal environment, the title of a Seminar at the University of Siena emerges from the web [2]: "La "centralità" della legge e la letteratura grigia. Profili di politica del diritto in Italia tra Otto e Novecento"[2]. From these findings, the idea of a survey on the wide variety of grey material available on Italian web portals arose. A first analysis showed that this material is available in different forms and dissemination is carried out through various means such as thematic bibliographies, newspapers articles, various types of documents published in PDF format or simple descriptions on web sites.
The following are a few examples excerpted from some home pages belonging to our corpus:

- ✓ Collection of *grey literature*. The Historical Archive of Women candidates for becoming the repository where the memories about these themes will be stored…[3]
- ✓ *Grey literature* ... Master copies · Reprints · Unreleased copies · *Grey Literature*. Archives for the history of Education…[4]
- ✓ *Grey literature*. The high quality brand of parks…[5]
- ✓ … International *grey literature*; national and international legislative data on the topic of drug addiction and related themes; documentary archive ...[6]
- ✓ Besides literature in German, there is literature in other languages and *grey literature* as well – in particular catalogues of museums and exhibitions…[7]
- ✓ It is about a few thousand of books, brochures and documents of *grey literature* concerning two topics, Italian dialects and minority languages…[8]

Given this scenario, this research aims at verifying whether – and eventually how much – the grey literature available on the web is actually structured, accessible or even managed by systems dealing with its organization and aiming at its retrieval and storing. The utmost goal is to build up a map of non-academic communities and their mechanisms for managing, presenting and disseminating this type of material. It is a sort of journey among the streams of the Web, which channel meeting minutes, manifests, fliers, pictures, newspapers articles, journalistic services and audio/video material on various topics. These "grey" products – by conveying basic information about social and popular culture – store, represent and spread knowledge.

[1] Famous linguist and former Italian Ministry of Education.
[2] The Seminar presents the results of a 2007 PRIN research project funded by the Italian Minister of Research: «Perpetue appendici e codicilli alle leggi italiane». Le circolari ministeriali, il potere regolamentare e la politica del diritto in Italia tra Otto e Novecento.
[3] www.archiviodonne.bz.it/it/progetti/raccolta-di-letteratura-grigia?
[4] www.historied.net/portal/index.php?option=com_content...id...?
[5] http://ambiente.regione.emilia-romagna.it/parchi-natura2000/consultazione/pubblicazioni/letteratura-grigia
[6] http://www.ceisroma.it/upgrade/biblioteca-agora/
[7] http://artlibraries.net/allg_infos_it/KUNST_GDK.php
[8] http://www.reteitalianaculturapopolare.org/archivio/fondi-in-rete/fondo-tullio-de-mauro.html

2. Materials and Methods

In order to get a specific range of data, we selected a sample data from the Italian Web portals using the following query criteria:

1. Search for >"letteratura grigia" OR "letteratura non convenzionale" OR "documentazione grigia" OR "materiale grigio".
2. Creation of the corpus
3. Analysis of the communities and the grey material retrieved
4. Statistical processing of data

We gathered the

- ✓ Type of Communities and stakeholders
- ✓ Type of production
- ✓ Terminology

and analysed the documents by

- ✓ a) Type of products
- ✓ b) Format
- ✓ c) Availability/Access

We got about 42,000 indexed documents considered as "interesting" with respect to the terms of the query. We have to consider that the list ordered by relevance of URL provided by Google never exceeds 1000 results. In order to have the largest number of results, we used the option proposed by Google about the data visualisation…

…In order to show you the most relevant results, we have omitted some entries very similar to the 390 already displayed. If you like, you can repeat the search with the omitted results included.

The first step was to "skim" the information retrieved by removing the duplicates and the irrelevant information. We obtained results consisting of 800 URLs for building the sample data (Corpus). Data was collected between May and July 2014. Then we extracted the information and built the Corpus by grouping the gathered data in 12 informative classes (Categories) assigned to each URL:

Stakeholder, *Type of Stakeholder*, *Name of Stakeholder*, *Infrastructure/service*, *Field*, *Topic*, *Type of material*, *Format*, *Access*, *Domain*, *Description*, *Terminology*. Finally, an analysis of the results followed.

3. Results

3.1 Stakeholders

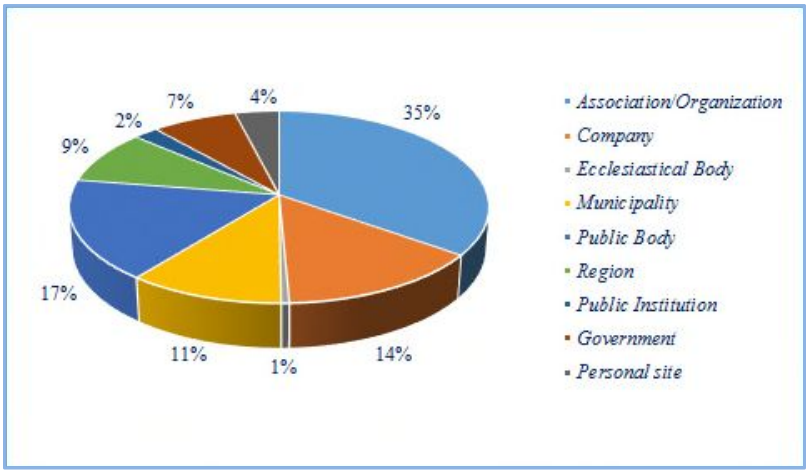
The institutional nature of the analyzed stakeholders is either public or private (Graph 1); in addition to a certain amount of personal sites, the highest percentage is represented by websites of associations/organizations as well as public bodies/institutions and companies which indeed have both public and private nature.

In the public sector, the institutional websites are those prevailing: they are those of Ministries, schools, public bodies, local or territorial authorities, healthcare agencies (in Italian, "Azienda Sanitaria" - ASL) whose websites contain detailed information about the structure and the services offered.

There are also thematic sites carried out by several collaborating administrations with a view to a precise purpose such as the presentation of a project or event, the provision of a service or the promotion of a specific field: an example is the portal called "Innovatori PA: la rete per l'innovazione nella Pubblica Amministrazione Italiana" where the title *Letteratura grigia e dati: accessibili, intellegibili e riutilizzabili* (about the advertisement of a 2014 workshop) can be found.

This website does not contain collections of grey literature but aims at disseminating the digital culture pertaining to the grey material, as for instance the sharing of databases for the complete digitalization of Italian public administration[9]. Amongst the associations/organizations there are professional, healthcare and research associations, Foundations and Consortia; to a small extent, also trade unions and socio-cultural and political groups are represented.

[9] Please visit the portal of the "Agency for Digital Italy" of the Presidency of the Council of Ministers (http://www.agid.gov.it/dati-pubblici-condivisione/basi-dati-pa)



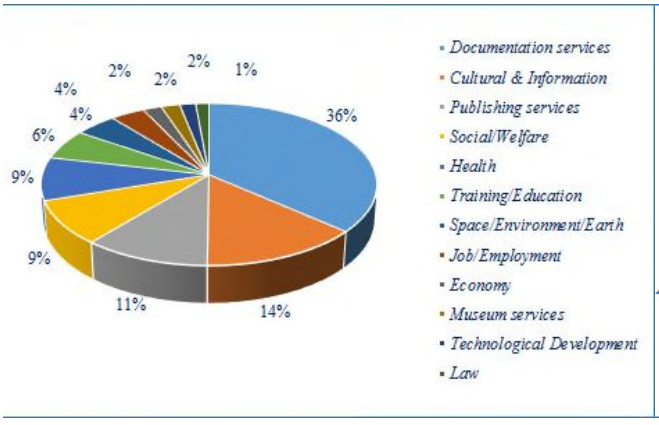
Graph 1 - Stakeholders

In the private sector, there are commercial companies, software houses, publishing houses, banks and management consulting companies, together with personal homepages, fora and blogs dealing with different topics which might have a social relevance such as alcoholism, obesity and violence against women. These stakeholders often restrict themselves to simply present grey literature as citations only but sometimes they do provide interesting editorials and grey literature of various nature.

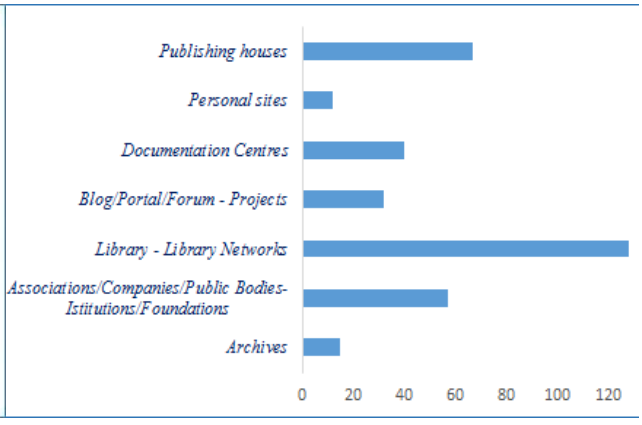
3.2 Thematic areas

Graph 2 shows the fields of application of the stakeholders: documentation services (36%) and cultural/information services (14%) are the areas where grey literature is more easily retrievable compared with fields as Training/Education or Job/Employment which are far more specific. Moreover, the management of grey literature is traditionally assigned to documentation and information services and this area actually embraces a wide variety of sub-fields: from technological development to e-commerce and services for online degree thesis, as well as services for online dictionaries and encyclopedias.

In most of the cases the infrastructure through which the producer manages the grey material is easy to identify: Graph 3 unequivocally shows that in most of the cases libraries and library networks keep the documentation though also archives and documentation centers are sometimes involved in handling this material. It frequently happens that also publishing facilities of associations, companies or institutions maintain this type of material.



Graph 2 – Fields of application



Graph 3 - Infrastructures

In some cases the identity of the supporting infrastructure was not specified and then we assumed that the stakeholder itself was taking care of the grey material, sometimes by creating ad-hoc fora, blogs or portals.

3.3 Communities

Our attention moved towards communities particularly up-to-date and socially relevant. Significant examples have been identified on websites presenting a heterogeneous set of material, which could prove to be especially interesting to researchers, scientists, professionals and simple fans of the various subjects if ever made available and usable.

Several communities handling and disclosing "grey" documentation in various ways come to light from this analysis. The ensemble of information collected allowed to identify and delimit four illustrative communities presenting the following matters: history of women's culture and of their movement in some specific Italian regions [3]; projects dealing with the sustainability of urban environment with respect to childhood and adolescence [4]; parks and other natural protected environments [5]; archaeological documentation such as draft reports, diaries from the site, letters and miscellaneous documentation [6] [7]; nursing and health-related disciplines which produce guidelines, diagnostic and therapeutic courses, informative material for patients and their families [8]. Also the theatrical culture is nourished by "grey products" as video archives, collections of music LPs and CDs, brochures, scripts, autograph manuscripts (i.e.. letters, correspondences, fliers, musical scores) [9] [10].

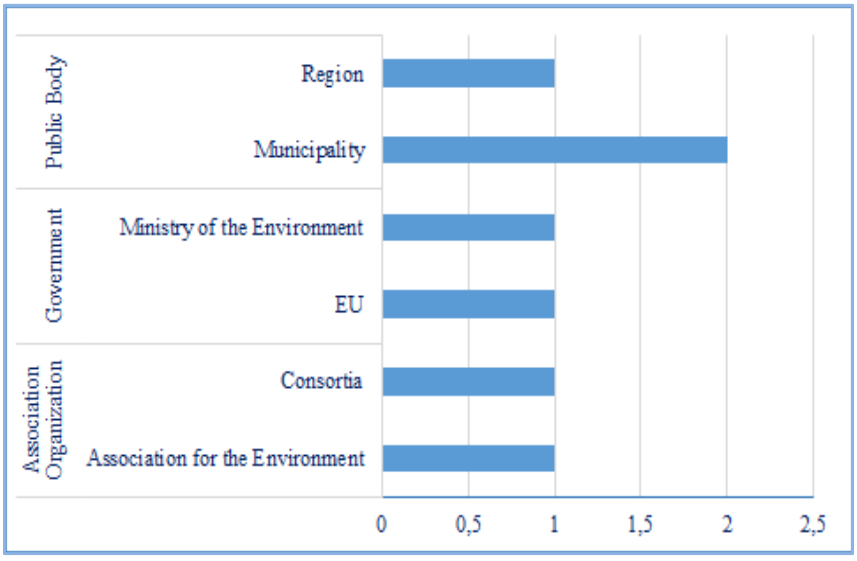
Finally we chose the following communities as case studies of this work: Environment, Women, Healthcare and History.

3.3.1 Environment

Environmental conservation is one of the main issues of today: and many governmental institutions and associations are in charge of preserving the national environmental heritage. Their nature is mostly public, in fact these are several municipalities that take care of managing documentation.

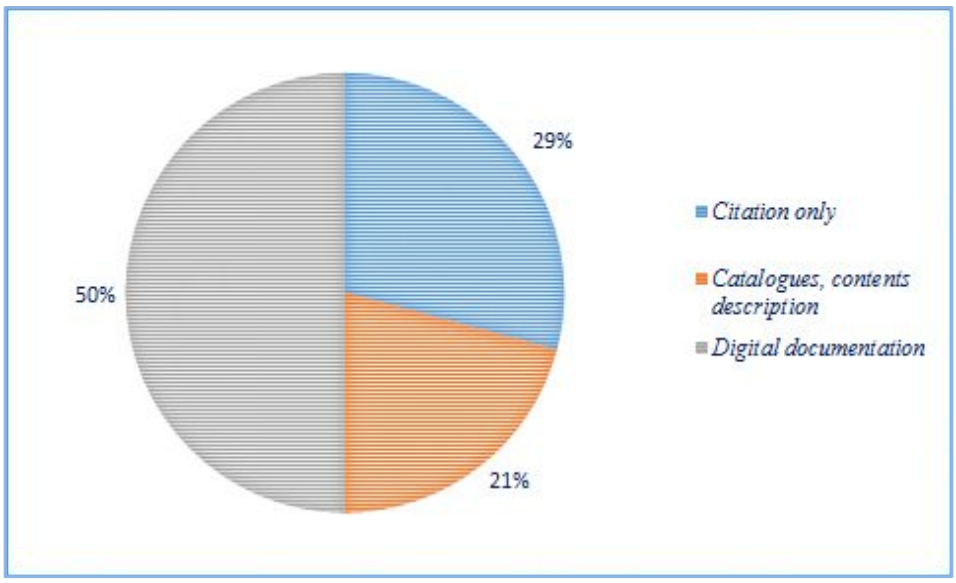
This community deals with the promotion of citizen's environmental awareness and solicits the adoption of adequate politics directly involving the citizenship in the care of the territory.

Several contents related to the environment can be retrieved from the web by means of grey material such as maps, data on city and marine pollution as well as thermal data collection.



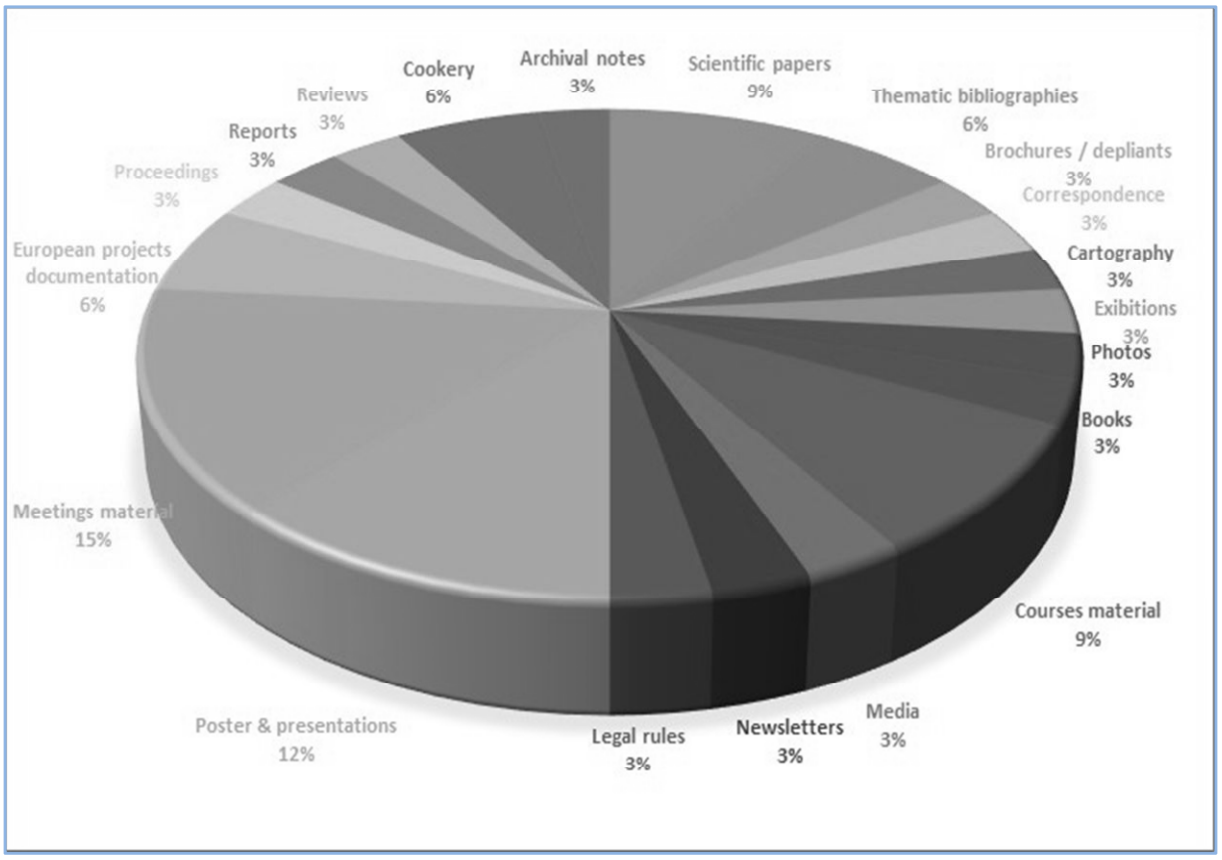
Graph 4 – Environment community – Nature of Stakeholders

Just to make a few examples, there are websites illustrating interesting projects of eco-sustainable cities for enhancing the quality of life, others – mostly institutional - endorsing some specific geographic areas where peculiar activities are carried on (for instance, mussel farming in the Puglia region in southern Italy) [11].



Graph 5 – Environment community – Document expression

Grey documentation of the environment community is varied: material related to meetings and events (fliers, brochures, etc.) and posters/presentations has the higher percentage but collections of photos and videos as well as thematic bibliographies are remarkable too.
Only in the 21% of the cases the community manages its contents by means of a structured database with search and information retrieval features, therefore an enhancement of content management and then fruition would be highly beneficial.

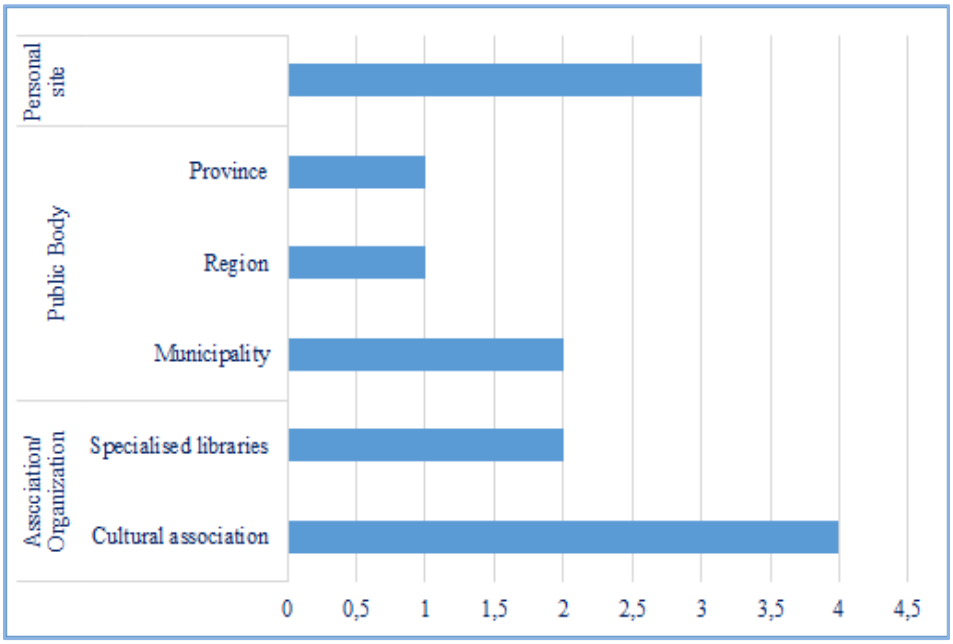


Graph 6 – Environment community – Type of contents

The material is freely accessible and the common format is html (70%) but in several occasions html is used in conjunction with the PDF format so that the latter is used in the 50% of the cases.

3.3.2 Women

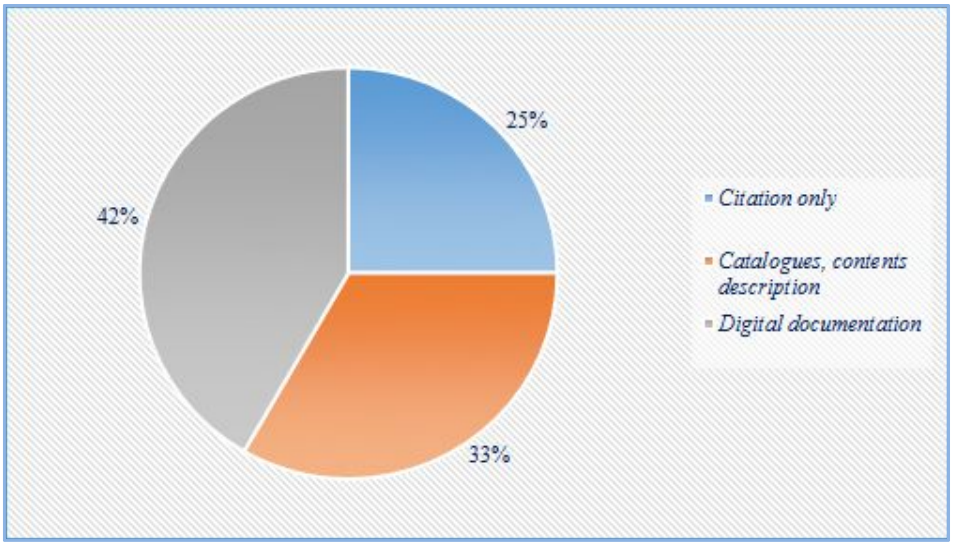
There is a wide documentation about problems related to women on Italian web sites of associations and institutions dedicated to these topics.



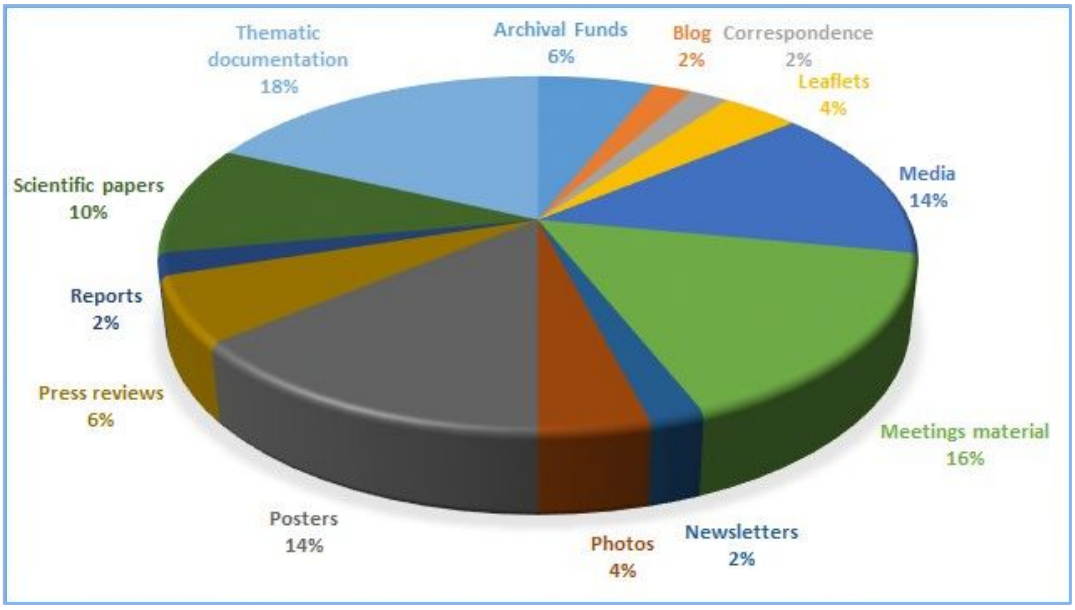
Graph 7 – Women community – Nature of Stakeholders

History of women is disseminated thanks to grey material stored in archives but also by means of websites which publish literature of various ages pertaining with the historical, cultural, political and social memory of Italian women and of their feminist movement for emancipation.

A good example is provided by the "Biblioteca Unione Femminile Nazionale" [12], a library established by the Milanese suffragettes more than a century ago and still offering nowadays a wide catalogue on feminine literature. This community usually pays more attention than the other three to the organization of the grey material in databases/catalogues (33% of documents); in the 42% of the cases, documentation is digitalized and is about the historical, cultural, political and social memory of Italian women: it is worthy mentioning the material on the feminine movements for emancipation of the XIXth century – first half of XXth century.



Graph 8 – Women community – Document expression



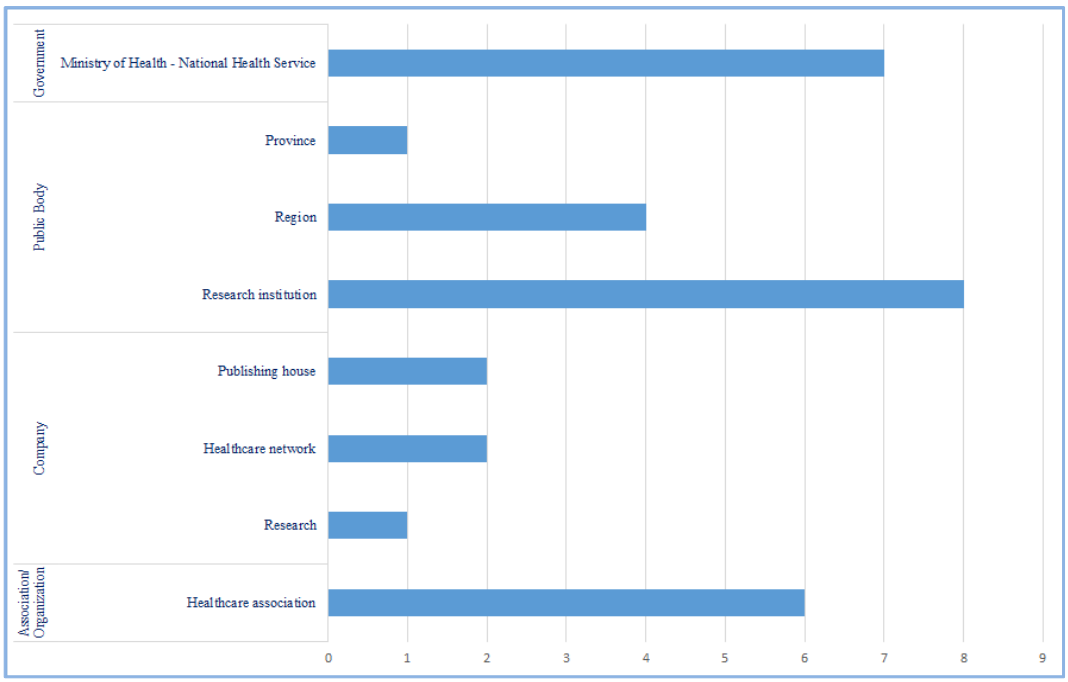
Graph 9 – Women community – Type of contents

As for the type of contents, thematic documentation focusing on women's rights, thoughts, wishes and creativity stands out: on these websites – alongside with worrying statistical data on feminicide and violence against women - there are maps locating the national and foreign Centers dealing indeed with violence against women. Then it comes various material on every kind of meetings, events, manifestations organized on this theme.
A good percentage is represented by the section called "Media" containing collections of movies, audio-visuals and photographs.
Also in this case the material is freely available and the most common format is html (85%) while the use of PDF is around 45%.

3.3.3 Healthcare

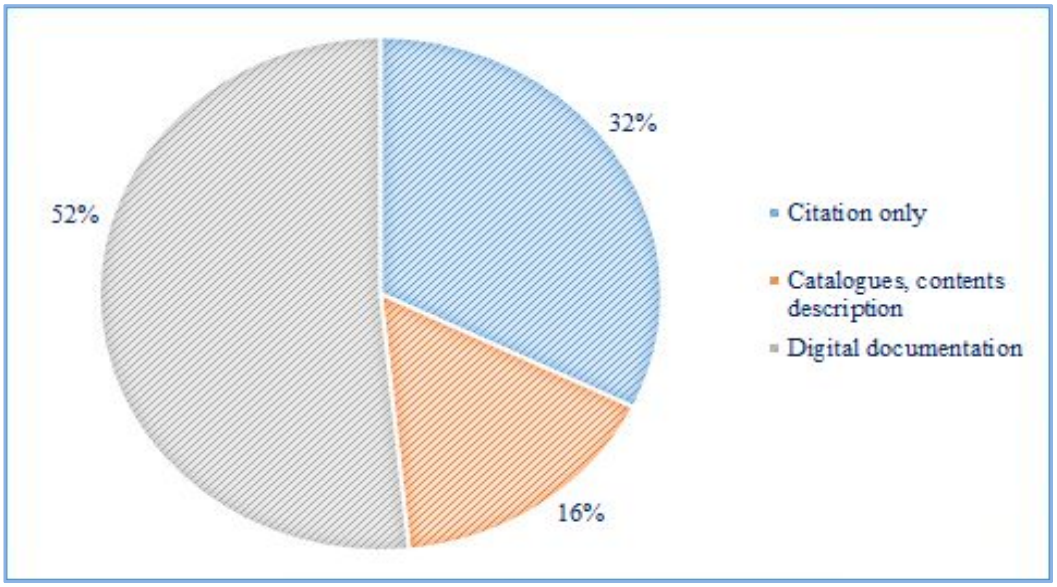
The Italian healthcare community is very rich in terms of initiatives and material available: the Ministry of Healthcare [13] is the central body of the National Health Service and is dealing with the safeguard of public health. This means that the main stakeholders of this community are mainly governmental and research organizations - both at central and outlying level - but there are also several public and private associations or institutions which devote ever more web pages to the fight against alcoholism, drug dependency and to the prevention of severe diseases (e.g. obesity). A significant example is represented by the "Istituto Ortopedico Rizzoli"[10] which is carrying on a project of an open archive of the various material produced within the Emilia-Romagna healthcare agencies, paying special attention to non-conventional literature while recalling the importance of knowledge and dissemination both in terms of diagnosis and therapy [14].

[10] http://www.ior.it/



Graph 10 – Healthcare community – Nature of Stakeholders

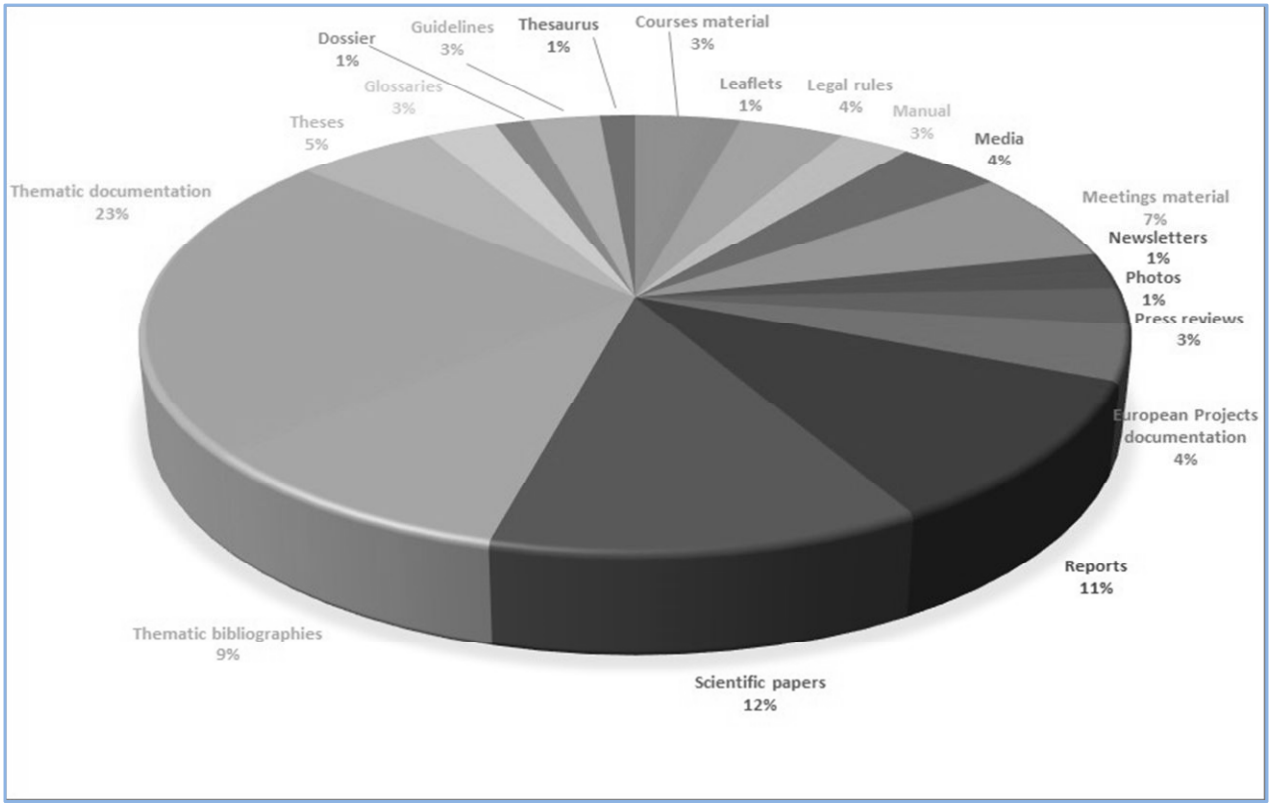
Other issues, amongst the most studied, concern mental health affections and youth problems, health and safety at work; furthermore, it can be noted the slow appearance of websites where grey material about complementary and alternative medicine (*CAM*)[11] starts being collected (thought not in a very incisive way yet). Although within this community the percentage of stakeholders using ad-hoc instruments to provide access to the documentation is quite low (about 10%), interesting initiatives are detectable, such as the Archives of grey literature of the "Oncological Center" of Aviano [15] which collect educational material related to training events, protocols and guidelines for welfare activities, references to degree theses, surveys and research works, working manuals and any other document of public interest produced by the Center. Access to consultation of the database is free as well as the downloading of documents, both in html (50%) and in PDF format (50%).



Graph 11 – Healthcare community – Document expression

[11] http://www.treccani.it/enciclopedia/dalla-medicina-alternativa-alla-medicina-integrata_%28XXI-Secolo%29/

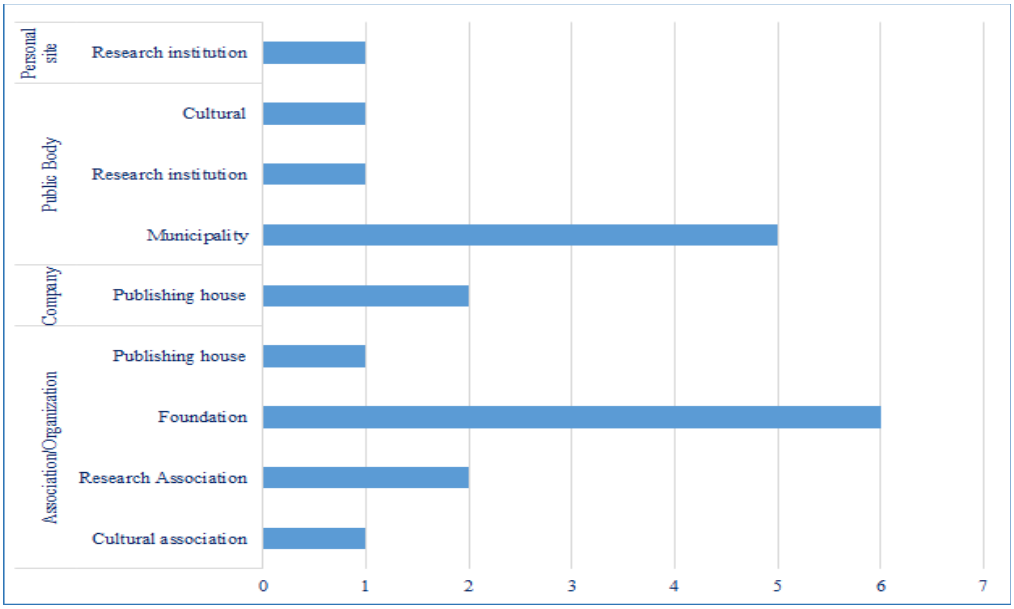
Graph 2 describes the partitioning of documentation within the healthcare community, highlighting how the thematic documentation definitely prevails but also pointing out how much the scientific production (papers, reports, etc.) is relevant in this field; on the other hand, by comparison with the environment and women communities, a smaller percentage is represented by conference, workshop and tutorial material.



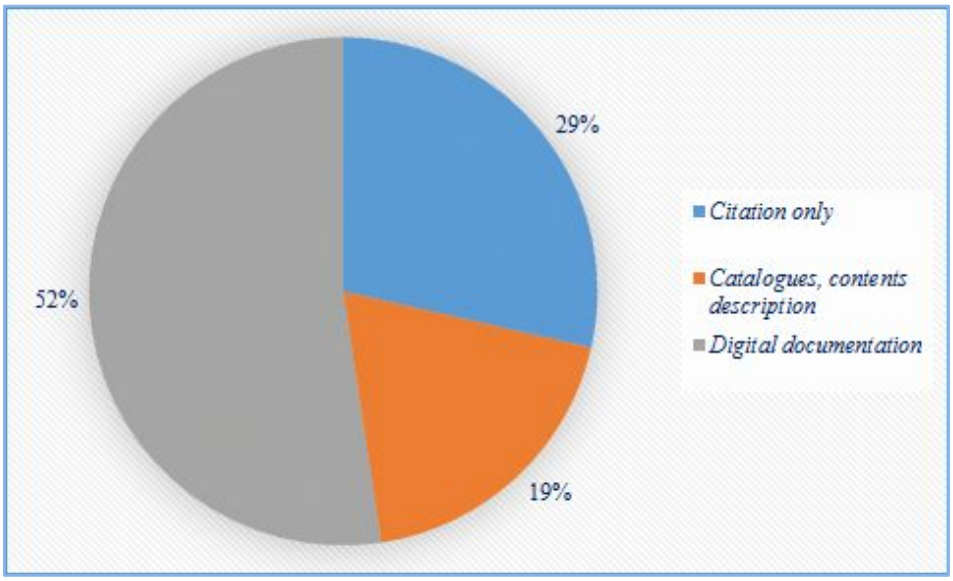
Graph 12 – Health community – Types of documents

3.3.4 History

This community is obviously very rich in terms of documentation: there are hints of local history provided by institutional sites (municipalities and provinces) and usually concerns territorial geography and customs; Foundations are nevertheless the major producers of grey literature of this community and a good percentage is also represented by research associations.

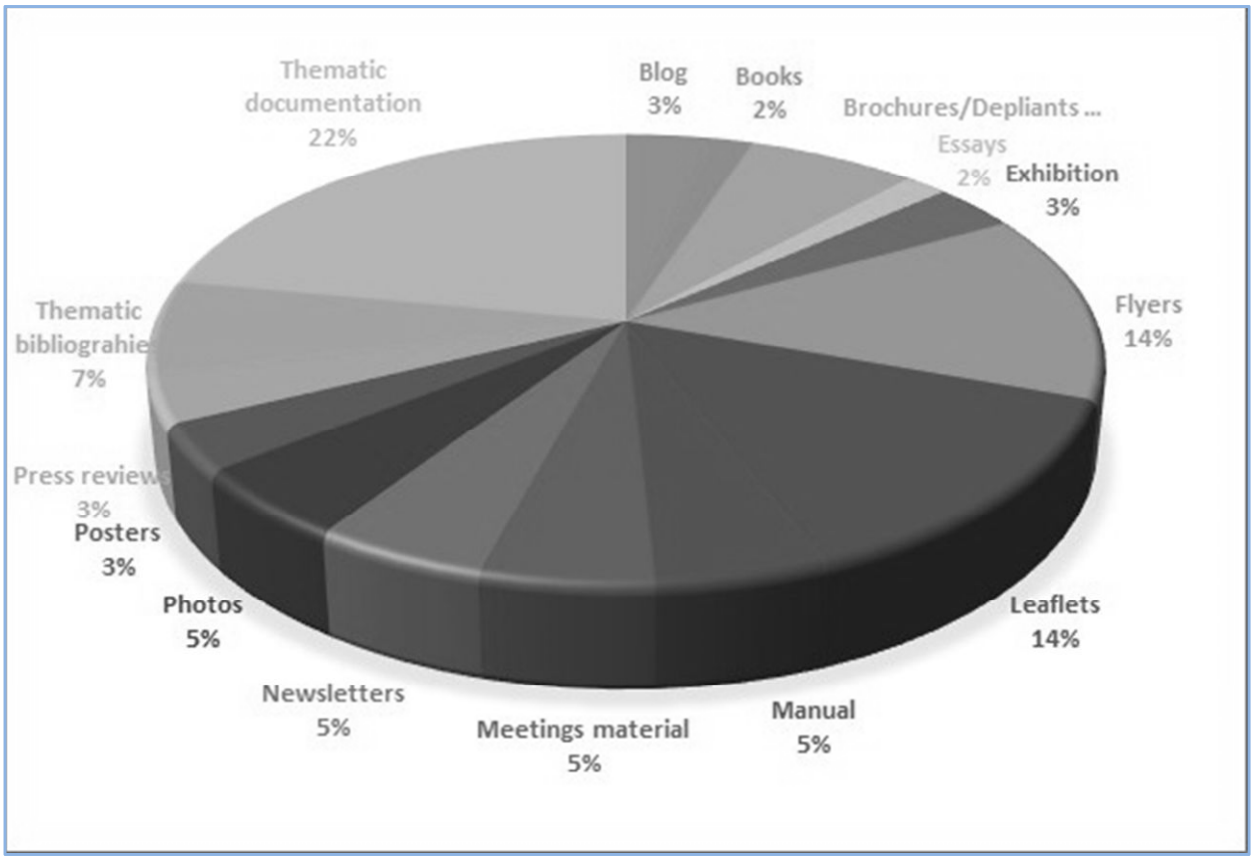


Graph 13 – History community – Nature of Stakeholders



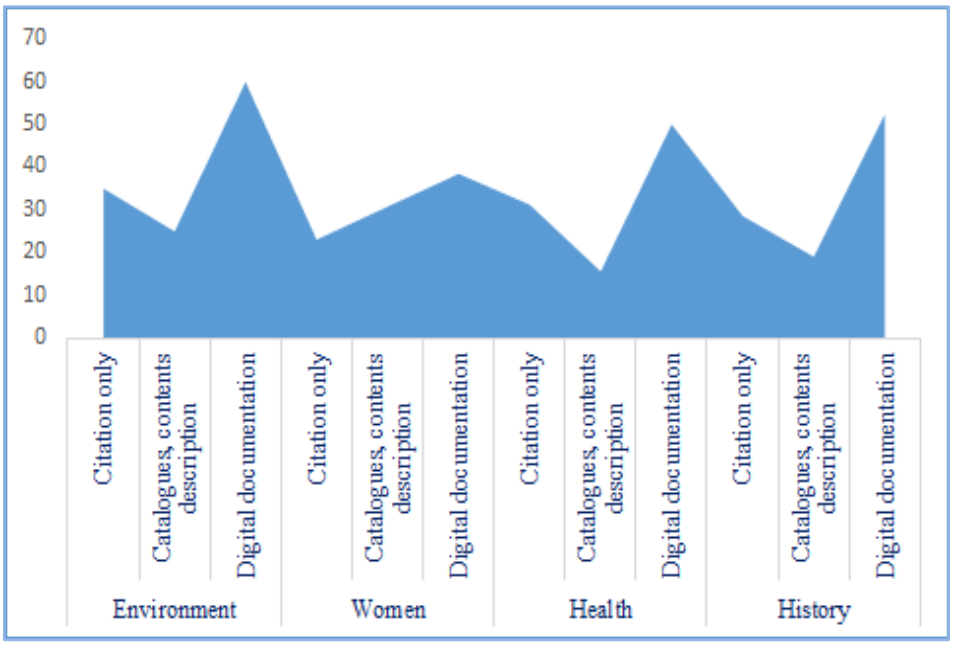
Graph 14 – History community – Document expression

Of course information on national history disseminated by institutes such as the "Istituto di Storia del Risorgimento Italiano" [16] can be easily retrieved: in these cases, grey material reflects refresher courses or more specific activities for both lecturers and students and is mainly made of miscellaneous documents – historiographical, archival, literary, films and direct evidences.
There are also websites devoted to the history of political movements or of single politicians which collected a huge quantity of material such as correspondences, photographs, cuts out, leaflets, etc., which are made available on the Web for avoiding the dispersal of significant pieces of the Italian history.



Graph 15 – History community – Type of contents

The material produced by this community largely belongs to the category of thematic documentation, followed by a good amount of documents like leaflets and flyers; there is also the typology 'essays', completely absent in the others. The contents available are offered completely for free, but the percentage of documents downloadable in PDF is lower than in other communities - around 30% - compared to html format (70%).
Within this community, we would like to illustrate an example of "work in progress" such as the "Galileo Galilei Foundation" [17] of our own town, Pisa. The Foundation takes care of disseminating scientific culture by promoting and enriching the heritage of the "Museum of Computing Instruments" which represents a kind of journey through the history of calculation along at least 500 years, from the calipers to the most modern computers. This newly established Foundation has a wealth of very important grey literature consisting of: manuscripts and unpublished correspondence of the major names in the history of science; material related to exhibitions and conferences; educational initiatives and guided tours for schools, operating manuals, etc. Unfortunately this amount of documentation is not accessible because has not been catalogued yet.
Graph 16 shows the progress of the different forms of expression of the documentation proposed by the four Communities. For all of them the highest percentage is the one that aims at the usability of content, compared to the citation only or the bibliographic description and/or cataloguing.
The Environment community offers the greatest amount of content, while the smaller amount belong to the Women community.



Graph 16 – Document expression

Rather equivalent seems to be the number of producers who only cite the presence of grey documents, while the highest percentage in the creation of bibliographic descriptions and/or cataloguing belong to the community of Women.

4. Multi-grey-expressions

From a terminological point of view, the expression *Letteratura Grigia* is the mostly adopted (86%) on the websites of this survey; the other lexical forms used in the query are nevertheless present: the term *materiale grigio* is used in the 12% of the cases only, *documentazione grigia* even less (2% of the cases) while *letteratura non convenzionale* can be found only on two websites.
Quite often the term "letteratura grigia" (and its variants) is followed by the explanations of its meaning as if the producers feel the need to explain what this "niche" term really stands for with words easily understandable to non-experts as well. A few examples excerpted from Italian websites:

- ✓ *…letteratura non convenzionale; quella vasta porzione di preziosi documenti che esula dal circuito commerciale...;*
- ✓ *…detta letteratura grigia scientifica non verificata, mentre quella controllata verrà successivamente pubblicata nelle riviste …;*
- ✓ *…"letteratura grigia" che, per le sua stessa natura, risulta difficilmente reperibile ed accessibile; questa parte del fondo, costituita prevalentemente dal cosiddetto "materiale grigio" (volantini, opuscoli, bollettini, ma anche appunti…;*
- ✓ *…letteratura non convenzionale (materiale grigio)…;*
- ✓ *…nella biblioteca si conserva anche il cosiddetto "materiale grigio", cioè tutte le informazioni su eventi relativi a spettacoli teatrali, cinematografici, convegni…;*
- ✓ *…con l'espressione letteratura grigia si indicano i documenti che non sono stati pubblicati da un editore, ma che possono costituire materiale di riferimento…;*
- ✓ *…materiale grigio (documenti non editi, materiale di corsi di formazione, ricerche della scuola…)...;*
- ✓ *…frutto dell'attività culturale e didattica prodotti nel corso degli anni da studenti e professori (materiale grigio) e materiali in forma multimediale e audiovisiva…;*
- ✓ *…il termine "letteratura grigia" indica la letteratura ed il materiale (opuscoli di organizzazioni, associazioni, ditte, unioni, relazioni su congressi e convegni)…;*
- ✓ *…il rischio sarebbe la scomparsa della cosiddetta 'letteratura grigia', ricerca pubblicata in forma non commerciale, spesso in copia unica, che...;*
- ✓ *…letteratura grigia … quella galassia della produzione scientifica che precede l'offerta dei risultati di una ricerca al peer review…;*
- ✓ *…si va così creando una particolare forma di letteratura grigia digitale, non nuova in termini assoluti, essendo l'erede degli archivi cartacei di «pre-print»..*

5. Conclusions & Remarks

First of all, we have to admit that the processing of data has been time-consuming and little fulfilling for the following reasons: i) there is a big amount of disorganized and badly-structured material; ii) there are many web pages with access denied; iii) the term "letteratura grigia" often appears simply as a description of the material but the content is not always available (e.g. catalogues, lists of references, glossaries, encyclopaedia entries, etc.).

The use of Google as a source for building our grey literature corpus has been useful for evaluating the state-of-the-art and for starting sketching a map of communities and stakeholders outside the academic scenario. However, using Google also presents some disadvantages: i) there is an "explosion" of information and is often difficult to discern who is producing it (a single person, a society, a municipality, a foundation). ii) the pages might contain a large amount of non-pertinent references to grey material (e.g. grey stones, grey furniture and so on); iii) many websites cannot be accessed, change their URLs or are not updated and the information is either cancelled or duplicated; iv) it is impossible to have more than 1000 results.

What is evident is the great variety of non-academic communities, each of them having a different approach to grey material; grey documentation is definitely ubiquitous and belongs to the most diverse fields but at the same time is more fragmented and variously represented than conventional.

The Web surely helps in disclosing what is available as well as in pointing out the difficulties which the communities still face in organizing the grey material: repositories dedicated to grey literature are inadequate and the shared feeling is the lack of willingness or possibility to build platforms for storing great quantity of grey material and equipped with engines for querying and retrieving information as well as semantic search engines for channeling information to the user: documentary fragmentation usually means loss of knowledge for the citizens.

However, we should admit that beyond the Academy, the Italian scenario of grey literature is rather rich and despite the lack of a methodical processing of the documentation, the awareness of the importance of grey literature and its dissemination comes to light.

References

[1] <http://www.reteitalianaculturapopolare.org/it/archivio/fondi-in-rete/fondo-tullio-de-mauro.html>

[2] <http://www3.unisi.it/eventi/prin2007/index.htm>

[3] <http://www.archiviodonne.bz.it/it/progetti/raccolta-di-letteratura-grigia>

[4] <http://www.cittasostenibili.minori.it/bibliografia/grigia.htm>

[5] <http://ambiente.regione.emilia-romagna.it/parchi-natura2000/consultazione/pubblicazioni/letteratura-grigia>

[6] <http://www.openpompei.it/2014/01/31/open-archaeological-data-for-dummies/>

[7] <http://generazionediarcheologi.myblog.it/2013/01/30/gli-open-data-in-archeologia-chi-li-conosce-chi-li-usa/>

[8] <http://www.news-medical.net/news/2007/04/19/17/Italian.aspx>

[9] <http://www.archive.org/movies/movies.php>

[10] <http://www.luccaterre.it/it/dettaglio/7423/Teatro-del-Giglio-Miranda.html>

[11] <http://www.sceglieremitili.eu/raccolta-di-fotografie-video-letteratura-grigia-e-scientifica-pertinenti-alla-filiera-mitilicola>

[12] http://www.unionefemminile.it/biblioteca/

[13] http://www.salute.gov.it/

[14] http://www.news-medical.net/news/2007/04/19/17/Italian.aspx

[15] http://www.cro.sanita.fvg.it/

[16] http://www.risorgimento.it/php/index.php

[17] http://www.fondazionegalileogalilei.it/index_en.htm

How Information in Grey Literature Informs Policy and Decision-Making: The Need to Understand the Processes

Bertrum H. MacDonald, James D. Ross, Suzuette S. Soomai, and Peter G. Wells
Dalhousie University, Canada

Abstract
Effective advocacy for grey literature must be based on understanding the environments in which it is used. As advances in communications technologies continue to occur at a seemingly breath-taking pace, all forms of information are being affected. Evolving publication practices are presenting new communication opportunities, in addition to disruptions of established patterns, as long-standing genres are being reshaped by powerful technological and societal changes. Disruptions can cause discomfort and anxiety, but opportunities to promote the value of particular information genres also arise. Grey literature, for example, continues to be produced in large quantities, which suggests that its importance in communication may be increasing rather than diminishing. Advocates of grey literature may believe this genre is undervalued or misunderstood, but lobbying for grey literature in the absence of understanding the contexts in which it is or can be used will likely fail unless information activity in those settings is understood. One prominent context encompasses public policy and decision-making where grey literature is often present but typically not noticed. Policy and decision-making are notably complex processes and increasing attention is being placed on developing an understanding of the research-policy interface and evidence-based policy making in particular. Conferences (e.g., Science Advice to Governments, Auckland, New Zealand, August 2014), evidence information services (e.g., one launched in the United Kingdom in 2014), research programs and institutes (e.g., Environmental Information: Use and Influence, Dalhousie University), and other initiatives emphasize the importance of understanding the relationship between research and policy, a sometimes contentious and even dysfunctional activity. Drawing on findings from research conducted within the Environmental Information: Use and Influence research program, which involves governmental, intergovernmental, and non-governmental organizations, we outline roles for grey literature in policy and decision-making contexts. We note, for example, types of grey literature used in these contexts, we identify preferences for specific features of useable information by managers and policy makers, and we outline pathways of research evidence, some of which is produced as grey literature. Information use is a non-trivial phenomenon that must be understood in advance of advocating the value of grey literature.

To be effective, advocacy for grey literature must be based on understanding the environments in which this literature is generated and used. A recent case about hydraulic fracturing, a controversial subject in North America and Europe (see for example, “Hydraulic Fracking,” 2014), illustrates this point. After months of public consultations, publication of numerous open access papers, and extensive study, an independent review panel on hydraulic fracturing (fracking) appointed by the Province of Nova Scotia published its final report on the subject in August 2014 (Wheeler et al., 2014). Five days after the release of the report, the provincial government announced that it would introduce legislation “to prohibit high volume hydraulic fracturing for onshore shale gas” (Government of Nova Scotia, 2014). In a front page news story two weeks after the government passed the legislation to enforce the ban, David Wheeler, chair of the review panel, stated categorically that the panel’s report was “very clear in not suggesting a moratorium” (McLeod, 2014). From Wheeler’s perspective the government’s opposition to fracking was “for the most part, not based on scientific evidence.” He claimed “science doesn’t cut it in the public-policy realm….Instead, a combination of social and political factors…come together to determine how a community assesses risk,” which “includes everything from knee-jerk fear to an understandable skepticism of the government’s track record of environmental management.” Wheeler believes that “these views can change with research and discussion” (McLeod, 2014).

The relevance of this incident to the topic at hand relates not to the subject of hydraulic fracturing *per se* but rather to the purpose of the review report, its audiences (both expected and unintended), and how its use can inform discussion about how to lobby for or highlight the significance of information published as grey literature. The final report by that independent panel,

and the ten discussion papers released earlier in 2014, are all examples of grey literature. Such documents are produced in the thousands by governments and their myriad agencies globally each year. The final report was designed primarily to inform the provincial government, although some might argue that politics rather than the evidence held sway in the government's decision announced very quickly after the report was released. The report served purposes other than simply to inform the government. Over several months, scientists, engineers, environmental groups, First Nations (aboriginal) groups, industry, and the public, principally the interested public, offered their views on the discussion papers, and media coverage increased awareness of the panel's work. Many people encountered and engaged with the grey literature produced by the panel. Moreover, as the documentation was made available primarily via an online website, residents and non-residents alike could, and did, stay abreast of the work of the panel and contributed to the panel's understanding of the science, engineering, economic potential, and public interest in the subject.

This Nova Scotian example demonstrates that information in grey literature and current information technologies facilitated an assessment and consultation process that is characteristic of modern policy making. Different types of grey literature were used and they served different purposes. However, it is likely that most who were involved with the process were unaware that grey literature fulfilled an important role in the consultation process, or that, in the absence of such material, the ability to become informed about the issues and participate in the debates would have been very difficult to achieve. Because of the generally transparent nature of the consultative process, made possible by social media, web-based technologies, and media coverage, the Minister of Energy in the provincial government was likely aware of public reactions and responses to the panel's consultation activities prior to receiving the final report. He would not have been aware of the panel's conclusions and recommendations prior to receiving the report, but he would have known about vocal opposition to fracking. While the government acknowledged receipt of the report, and in fact requested it, it is believed that advocacy for grey literature or the merits of the scientific advice it contained would not have given this report any more authority at the political level in this decision-making context. By identifying and understanding the variety of steps in this consultation and decision-making process, the place of information presented as grey literature and its use by numerous individuals can be appreciated. If advocacy for this literature is needed, then its promotion is likely to be more effective by elucidating the life cycle of such reports.

Advocates of grey literature may believe that this genre is undervalued or misunderstood, and evidence suggests that is the case (McKimmie & Szumak, 2002; Ravindranath, 2010; Thaler, 2010). Advocacy for grey literature can be affected by confusion about what actually constitutes this genre and whether it is reliable (credible). Grey literature experts may assume that the genre is clearly demarcated, but many information users do not recognize the distinction or consider it important.

Grey Literature in Policy and Decision-Making Contexts – Some Challenges

Grey literature is commonly used in policy and decision-making contexts, but it may not be noticed or recognized as grey literature. In such contexts, attributes of salience, credibility, and relevance can far outweigh peculiarities of definition (McNie, 2007). In a recent review, an author questioned why the reports produced by the Intergovernmental Panel on Climate Change were labelled as grey literature. After all, the reviewer pointed out, these reports "are probably some of the most thoroughly peer-reviewed publications in the world" (personal communication, October 2014). This reviewer simply did not view peer-reviewed technical reports as grey literature even though that individual acknowledged that the reports are "intergovernmental documents."

Within policy contexts, grey literature may be preferred and relied upon because its credibility can be confirmed, the relevance of the information it contains matches the expectations of readers, and the scale (local or regional versus national of global) fits the requirements for the decisions that need to be made. Moreover, the networks in which the policy specialists (managers and/or scientists) operate provide the assurance that the information is credible. These practitioners may actually view "grey" literature as not being credible because they assume it is not the type of documentation that they work with, even on a daily basis. In other words, the term "grey

literature" may unintentionally convey negative connotations in this context, a point that advocates for the genre must take into account.

Public policy and decision-making can have different manifestations according to particular political regimes or governance models, a point that merits emphasizing (Liverani, Hawkins, & Parkhurst, 2013). For example, recognizing the similarities and differences in the governance structures of countries, e.g., as regards public consultation, can be very important for understanding the role(s) grey literature fulfills and ultimately how the relevance and importance of this literature could be promoted.

Numerous studies (e.g., Lawrence, Houghton, Thomas & Weldon, 2014; Luzi, 2000; Schöpfel & Farace, 2010; Thelwall, Klitou, Verbeek, Stuart, & Vincent, 2010; Webster & Collins, 2005) have pointed to the growing production of information (often as grey literature) by government departments, international inter-governmental bodies, and non-governmental organizations, resulting in greater reliance on this publishing practice for disseminating information to assist with policy decisions. The quantity of information is now very large and the proportion that is published or otherwise distributed as grey literature is unknown. The number of scholarly documents available on the web, not including grey literature, was recently estimated at 114 million (Khabsa & Giles, 2014; see also Bornmann & Mutz, 2014 and Van Noorden, 2014). Another recent study estimated that a new journal article is published every 20 seconds (Munro, 2013). With the addition of grey literature the number of available publications increases markedly. As of five years ago, 95% of the web, composed of over 220 billion pages, was not indexed by search engines (Scheeran, 2012). The growth of the web over the past five years now makes it very difficult to estimate the extent of web-based information, which means that it may be impossible to determine the quantity of existing grey literature (BrightPlanet, 2014).

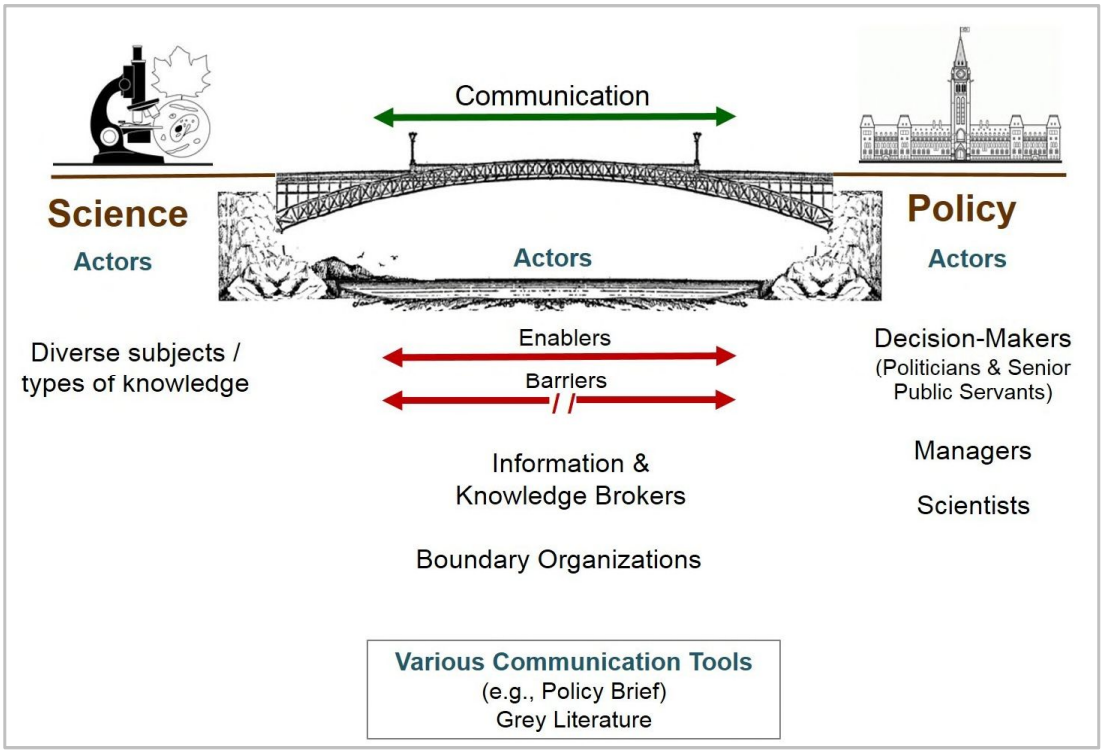
The massive volume of new web-based information has affected abilities of decision makers to discover relevant, timely information. What proportion of the grey literature component of the information universe is worth paying attention to and, therefore, merits efforts to increase its visibility to search engines by enhanced information architecture and embedded metadata? Examination of information activities in decision-making contexts, including how people deal with information overload (Bawden & Robinson, 2009; Parsons, 2010), could provide grey literature advocates with insights to address this question.

Science-Policy Interface(s)

Since 2002, the Environmental Information: Use and Influence (EIUI) research program at Dalhousie University has focused on understanding the use and influence of marine environmental and fisheries information and the processes at work at the science-policy interface (EIUI, 2014). Results from our recent studies support the view that an appreciation of the context(s) in which grey literature is used will inform efforts to recognize and promote its value.

Figure 1 provides a schematic illustration of our understanding of the science-policy interface(s). The "information universe" is multidimensional and information flow is often non-linear, leading to the complexity of activity at the science-policy interface. However, in its most simple form, the interface is characterized by information production (science) and information use (policy) with often bi-directional communication between these two broad categories. Furthermore, the science-policy interface is more inclusive than the label implies, as it encompasses many social processes. In fact, numerous interfaces exist due to many decision-making contexts in which issues, information types, actors, jurisdictions, and other factors vary (e.g., Sarkki et al., 2014). Actors or individuals at the interface compose both the "science" and "policy" arenas and they influence both the production of and preferences for information types, especially grey literature. Bridging these two are a suite of additional actors who serve in various communication roles, such as editors, publishers, and distributors, now functioning in mostly digital capacities. Numerous enablers and barriers to communication can influence information flow and in some instances a particular factor, e.g., the terminology used by research communities, may enable or impede communication.

Figure 1. Science-Policy Interface(s)



A diversity of subjects and types of knowledge contribute to the multiplicity of information products generated by the "science" component in Figure 1. Besides the natural and physical sciences, the social sciences and humanities, as well as traditional knowledge, also contribute significantly to the pool of information that can, and often should be, considered in decision making. Policies themselves can be administrative, legislative, or regulatory. Various types of individuals constitute the policy and decision-making bodies, e.g., politicians and senior public servants, among whom are managers and scientists who conduct research within the mandates of government departments and agencies. Furthermore, actors at the science-policy interface can include information and knowledge brokers whose synthesis and communication skills enable them to translate the information generated by either the science or policy communities into a form suitable for either one. Some organizations, such as non-governmental organizations, are particularly effective as boundary organizations at this naturally complex interface. All of the individuals who work and pursue careers in the various components of this interface operate within a diversity of world views, educational capacities, and constraints that affect the production, distribution, reception, and use of information. Communication can be affected, and often complicated, by many factors that converge here. As a consequence, various tools have evolved to promote communication among the array of actors fulfilling the roles found at this interface.

Measuring Use of Information Available as Grey Literature

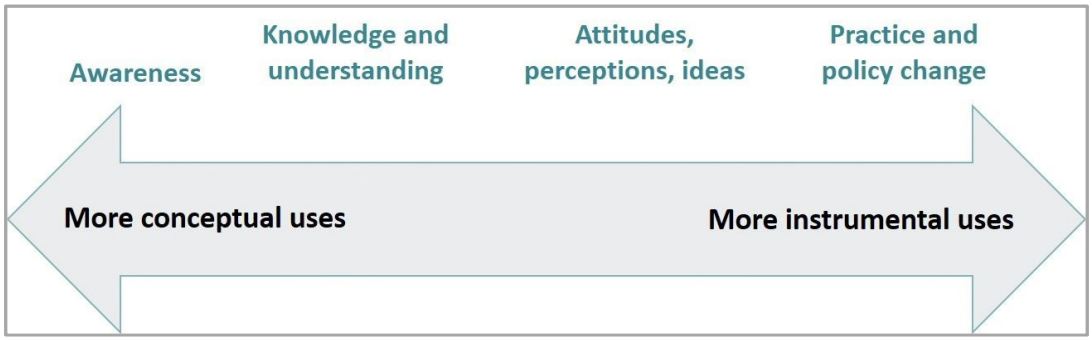
While grey literature is produced in large quantities by many organizations and is frequently used in decision-making contexts, the use of this large body of information in policy development is deemed extremely important but poorly understood (Ascher, Steelman, & Healy, 2010; Briggs & Knight, 2011; Holmes & Lock, 2010; Likens, 2010; McNie, 2007; Stojanovic, Ball, Ballinger, Lymbery, & Dodds, 2009). Few organizations have undertaken an analysis of the use of their publications, and information pathways in decision-making contexts are still being elucidated (Economic Commission for Europe, 2003; MacDonald, Cordes, & Wells, 2004; Soomai, MacDonald, & Wells, 2011a, 2011b; Wells, 2003; Wells, Duce, & Huber, 2002).

As intimated above, information use in policy and decision-making contexts is a complex phenomenon. It operates at various levels of scale: geographic, institutional, political, and temporal. Most models of scientific communication ignore the use of research information in public policy making, where information assimilation is different than in pure research contexts (Duff, 1997; Dunn, 2005; Doern, 2001; Søndergaard, Andersen, & Hjørland, 2003; Van der Veer

Martens & Goodrum, 2005). Decision makers or their advisors may find it challenging to choose which information to use when there is an absence of consensus among scientific experts and the presence of competing views advocated by stakeholders. In policy settings, information in grey literature may be given greater importance than that in peer-reviewed journals because the language is more accessible and a more rapid and flexible production cycle of grey literature can facilitate knowledge diffusion where decisions are based on competing factors, e.g., the pressures of political processes (Bremer & Glavovic, 2013; Shanley & López, 2009; Pielke, 2007). However, as open-access, peer-reviewed journals proliferate, papers published in these venues will be more easily accessible to decision-makers than has been the case previously, which may affect the mix of grey and primary literature used in decision-making contexts.

While information in grey literature may have a direct influence on the development of a policy, this type of "use" is only one of several. "Use" of information can, in fact, extend in meaning from general awareness through to increased understanding to a change in attitude about a subject or to actual implementation of the information in practice or policies (Figure 2) (Nutley, Walter & Davies, 2007; Young, 2014). This spectrum of use lays out a challenge to those wishing to promote the value and use of grey literature. An example drawn from one of our research projects illustrates this point.

Figure 2. Continuum of Information Use (Nutley, Walter, & Davies, 2007)



The Gulf of Maine Council on the Marine Environment is an American-Canadian intergovernmental organization consisting of representatives of three American states, two Canadian provinces, and the federal governments of both countries, as well as representatives from selected non-governmental organizations. This organization has produced several hundred publications over its history related to its mandate to maintain and enhance environmental quality in the Gulf of Maine (MacDonald, Cordes, &Wells, 2004). In one of our studies we interviewed members of the main Working Group of the Council about their understanding of the production and use of the Council's publications (Cossarini, MacDonald, & Wells, 2014). A member of that Working Group said the following about one of the publications:

We have a five or six page handout on [the] American Eel…and its status in the Gulf of Maine. So, I gave that to our pelagics advisor, fisheries advisor, and he found it very informative, a good synopsis and was quite impressed with it…now he never told me like that changed how [he] recommended to the Minister what our position is on it…but he found it informative in terms of getting his knowledge, in terms of playing that role as an advisor to the Minister of the issue.

This manager's description about the "use" of the fact sheet on the American Eel underscores the different aspects of information "use." Both the manager and the fisheries advisor were clearly aware of the fact sheet. That awareness led to greater understanding about this fish species; however, whether the increased understanding led to direct implementation of the information in policies revised or developed by the particular department of government is not revealed. Nonetheless, the manager's observations illustrate three types of information use. That particular manager's comments also demonstrate that advocates of the use of grey literature cannot assume that simply drawing such literature to the attention of potential users is sufficient. Each stage of use (awareness, understanding, change in attitude, and implementation) may require different promotional techniques, or a focus on different features of the information so that how it might be used becomes obvious to a target recipient.

Types or Formats of Grey Literature

We realize that experts on grey literature are well aware of the wide variety of types of information formats and products that constitutes this genre: internal policy documents, briefing documents, technical reports, contract or consultants' reports, internal reviews, fact sheets, infographics and posters, conference proceedings, newspapers, monographs, websites, social media posts, and a host of others. Acknowledgment that there is an array of types may lead to overlooking the role(s) that each fulfills and the need for different strategies for raising awareness and use of each type. Furthermore, given ongoing constraints on resource allocation and the vast quantity of grey literature, efforts to improve the use of grey literature will benefit from understanding the functions of the various types.

Briefing notes and memos, for example, are frequently used in public policy and decision-making contexts (see Figure 3). Policy briefs are frequently used to convey information to politicians and decision makers in many government organizations. This type of document, while very common, is often not readily accessible, due to both the confidential contexts in which it is used and the often ephemeral existence of the documents. Few studies have investigated the effectiveness of this form of communication (see Beynon, Gaarder, Chapoy, & Masset, 2012; Masset, Gaarder, Beynon, & Chapoy, 2013) and, interestingly, one recent study found little evidence that policy briefs actually led to a change in prior beliefs (Rajabi, 2012). In our view, the quantity and wide-spread use of this form of grey literature points to the importance of briefing notes and memos. In addition, our experience (SS Soomai and PG Wells) in government showed that briefing notes submitted to managers and decision-makers at various levels were regularly used as an effective way to distill information on a problem and offer options for solutions.

Figure 3. Selected Types of Grey Literature

Type	Role
Briefing Notes/Memos	Advice / The "Facts"
Technical Reports	Document Detailed Research Results
Social Media	Brief, Rapid, and Frequent Communication

Technical reports have been produced by governments and other organizations in very large numbers and in various formats and styles for decades. This type of grey literature has received extensive attention by grey literature experts, librarians, and authors alike. Often, the resources required to produce technical reports can be substantial. State of the environment reports, for example, may entail the commitment of many contributors, editors, illustrators, designers, and publishers, at considerable expense (Wells 2003). Enhancing awareness and use of these reports is merited on several grounds, one of which is accountability for fiscal resource allocation, including the costs associated with not preparing such environmental reports. Other types of technical reports may be far less demanding in production and consequently have little probability of use. As the roles for technical reports vary, the expectations about readership need to be taken into account (very few politicians have the time or the expertise to read lengthy technical reports, for example) and should have a direct impact on decisions about promoting this form of grey literature.

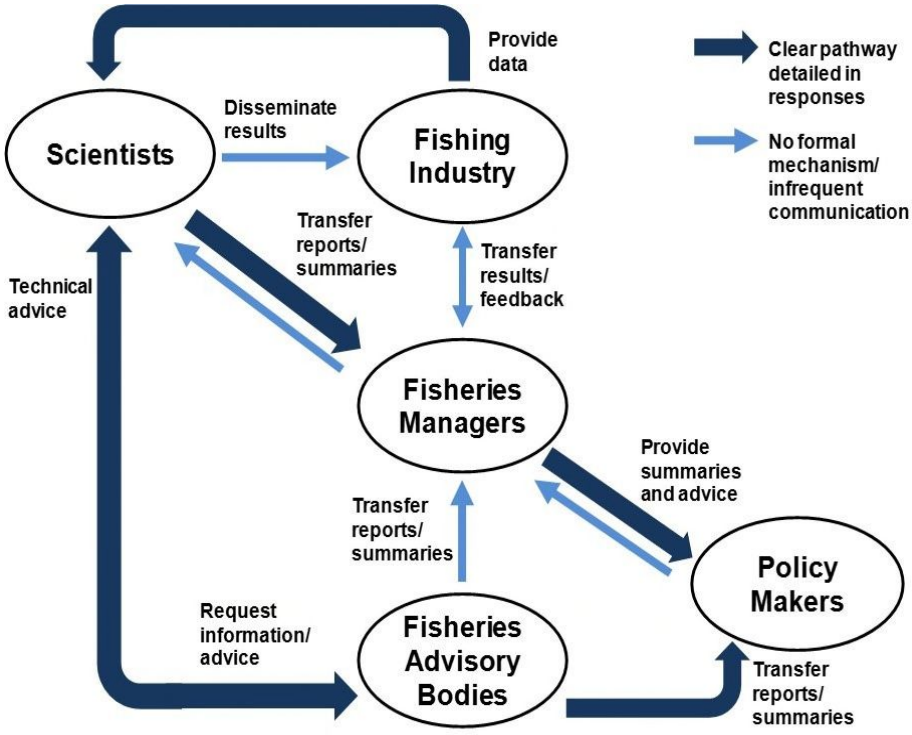
Rapid communication and rapid-fire feedback are not new to the research fields. But online publishing and social media are changing how people learn about new developments and how people incorporate this information into decision-making processes. How much researchers and

decision-makers engage in and monitor social media is also evolving. There is little doubt that this form of communication is taking on increasing importance and its role does merit careful consideration. However, the sheer volume of information flowing through social media presents considerable challenges for deciphering what portion is significant enough to warrant development of strategies for promoting its awareness and use.

Communication or Information Pathways

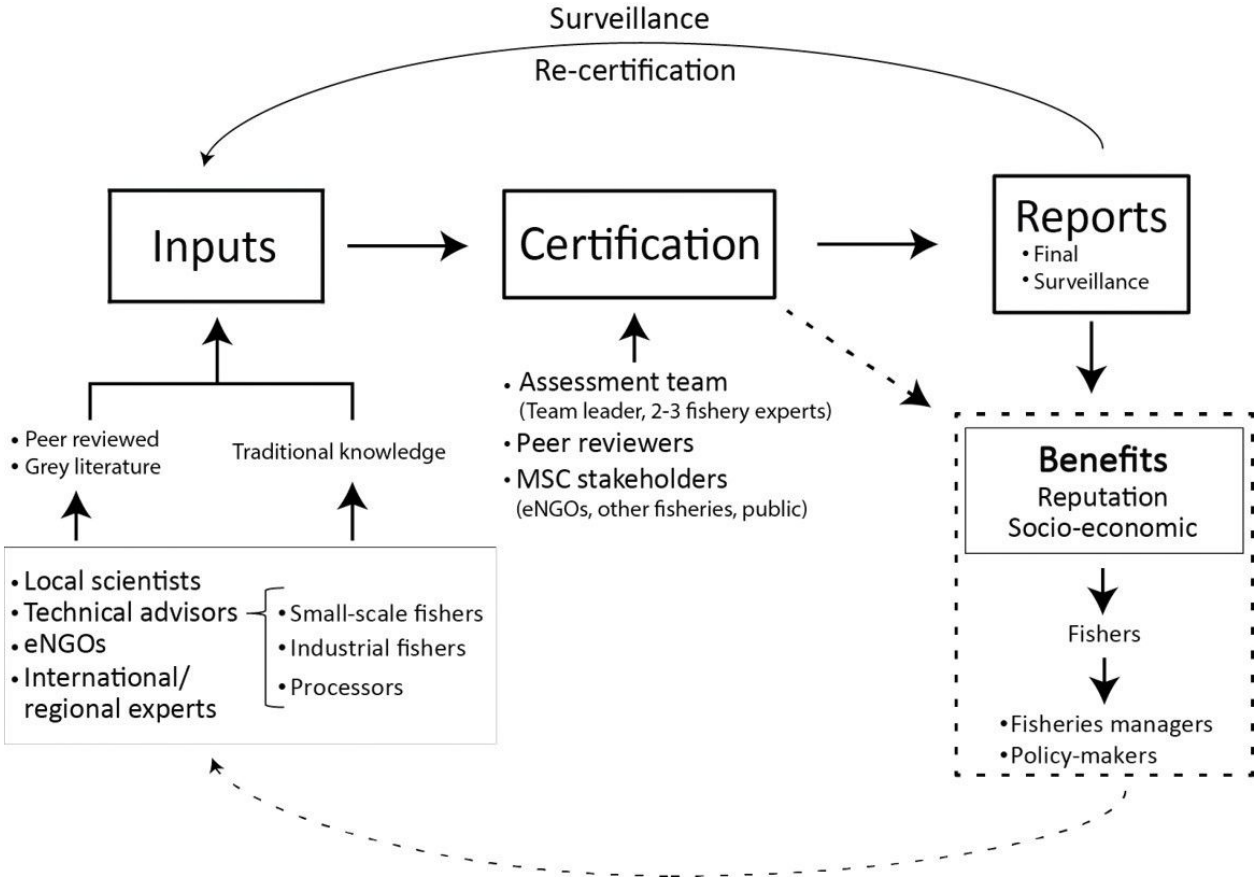
While understanding the types and roles of grey literature can inform advocacy for this genre, further appreciation of decision-making processes is needed to increase the likelihood that the advocacy will be relevant and effective. Communication patterns are changing due to the prominence of the internet and social media (Cossins, 2014), and other ever advancing information technologies, and this trend will likely continue. Non-governmental organizations have become more active at disseminating scientific information to policy communities at various levels of government (Grainger, 2013). However, "inherent differences between the fundamental structures and traditions of science and policy" (Francis, Whittaker, Shandas, Mills, & Graybill, 2005, p. 35) contribute to sub-optimal "flow of knowledge between researchers, policy makers, and resource managers" (Roux, Rogers, Briggs, Ashton, & Sergeant, 2006, p. 5). Therefore, tracing out information pathways that identify the points where grey literature plays a role is important. For example, information pathway models have informed our recent research (as seen in Avdic, 2013; Cano Chacón, 2013; Cossarini, MacDonald, & Wells, 2014; Hutton, 2009; Soomai, Wells, & MacDonald, 2011).

Figure 4. Information Flow: Fisheries Technical Reports (Soomai, Wells, & MacDonald, 2011)



In the first pathway model (Figure 4), the production and movement of information among five significant groups in fisheries activities in the Caribbean was studied, namely, industry, scientists, fisheries managers, advisory bodies, and policy makers (Soomai, Wells, & MacDonald, 2011). Identifying what information was assembled, produced, and used and in what format by each group provided an opportunity to determine where information flow and use were working well and where barriers to information exchange and use existed. This flow model equips an advocate for grey literature with understanding of where promotion of the value of this literature needs strengthening and also what information formats are preferred by each group. For example, while some members of the fishing industry can read and understand technical reports, most cannot. Thus, this model shows less use of technical reports by the fishing industry, and it is quite likely that efforts to promote use of this literature by this group would be unsuccessful.

Figure 5. Information Flow: Marine Stewardship Council (Cano Chacón, 2013)



In the second pathway model (Figure 5), the production and movement of information in the certification of sustainable fisheries by the international Marine Stewardship Council processes are set out (Cano Chacón, 2013). This process results in labelling consumer products when a fishery is determined to be sustainable. Grey literature is found at a number of points in the decision-making process. Learning that the major reports are typically only produced in English and made accessible primarily by posting to the Council's website presents an advocate for greater use of this grey literature with the understanding of how to tailor advocacy tactics. For example, simply drawing attention to the existence of the reports is not likely to increase use by small fishers in countries where the certification process is being applied.

Conclusions

As illustrated by the Nova Scotian example given earlier, information published as grey literature can fill an important role in policy-making contexts. Promotion of the value of such literature, in our view, is merited, particularly when it offers evidence that will help to resolve societal problems. In an age of information abundance, even information excess, many factors affect what if any information resources will be used to inform decision-making. As well, use itself is a complex phenomenon that is closely related to context. Evidence clearly competes with other factors that are brought to bear in decision making.

Lobbying for the value of grey literature should not begin with the assumption that the value is obvious to potential users. Rather, advocacy should begin from a position informed by an understanding of the range of settings in which it is used and with the various audiences or readers taken into account. The role that grey literature fulfils must be viewed within the larger background of information production and communication where the relative importance of information in grey literature versus information in primary literature can and will vary. While understanding the types and roles of grey literature can inform advocacy for this genre, appreciation of decision-making processes is needed to increase the likelihood that the advocacy will be meaningful and effective. Further research on information use is needed and with increased understanding resulting from such study, advocacy for grey literature will be better informed.

References

Ascher, W., Steelman, T., & Healy, R. (2010). *Knowledge and environmental policy*: *Re-imagining the boundaries of science and politics.* Cambridge: The MIT Press.

Avdic, V. (2013). Measuring use and influence: An assessment of FAO's flagship report *The State of world fisheries and acquaculture*. (Unpublished Master's thesis). Dalhousie University, Canada.

Bawden, D., & Robinson, L. (2009). The dark side of information overload, anxiety, and other paradoxes and pathologies. *Journal of Information Science, 35,* 180-191.

Beynon, P., Gaarder, M., Chapoy, C., & Masset, E. (2012). Passing on the hot potato: Lessons from a policy brief experiment. *IDS Bulletin, 43*(5), 65-75.

Bornmann, L., & Mutz, R. (2014). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology.* [Prepublication copy retrieved from http://arxiv.org/abs/1402.4578v2].

Bremer, S., & Glavovic, B. (2013). Mobilizing knowledge for coastal governance: Re-framing the science-policy interface for integrated coastal management. *Coastal Management, 41*, 39-56.

Briggs, S.V., & Knight, A.T. (2011). Science-policy interface: Scientific input limited. *Science, 333*(6043), 696-697.

BrightPlanet Corporation. (November 5, 2014). How big is the internet? *Deep Web University.* Retrieved from http://www.brightplanet.com/2014/11/big-internet/

Cano Chacón, M. (2013). The role of information of the Marine Stewardship Council certification process in developing countries: A case study of two MSC fisheries certified in Mexico. (Unpublished Master's thesis). Dalhousie University, Canada.

Cossarini, D. M., MacDonald, B. H., & Wells, P. G. (2014). Communicating marine environmental information to decision makers: Enablers and barriers to use of publications (grey literature) of the Gulf of Maine Council on the Marine Environment. *Ocean & Coastal Management, 96,* 163-172.

Cossins, D. (2014, October 1). Setting the record straight. *The Scientist.* Retrieved from http://www.the-scientist.com/?articles.view/articleNo/41056/title/Setting-the-Record-Straight/

Doern, G. B. (2001). Science and scientists in regulatory governance: A mezzo-level framework for analysis. *Science and Public Policy, 28,* 195-205.

Duff, A. S. (1997). Some post-war models of the information chain. *Journal of Librarianship and Information Science. 29*(4), 179-187.

Dunn, K. (2005). Impact of the inclusion of grey literature on the scholarly communication patterns of an interdisciplinary specialty. Paper presented at the Sixth International Greynet Conference, NYAM, December 2004. Amsterdam: TextRelease.

Economic Commission for Europe. (2003). *Environmental policy in transition: Ten years of UNECE environmental performance reviews.* New York: United Nations.

EIUI. (2014). Environmental Information: Use and Influence. Retrieved from www.eiui.ca.

Francis, T. B., Whittaker, K. A., Shandas, V., Mills, A. V., & Graybill, J. K. (2005). Incorporating science into the environmental policy process: A case study from Washington State. *Ecology and Society, 10,* 35. Retrieved from http://www.ecologyandsociety.org/vol10/iss1/art35

Government of Nova Scotia. (2014, September 3).Government to prohibit hydraulic fracturing. Retrieved from http://novascotia.ca/news/release/?id=20140903005

Grainger, R. (2013). How does information influence policy? The role of fishery organizations in policy-making for fisheries. Public Lecture, Dalhousie University, Halifax, 20 September.

Holmes, J., & Lock, J. (2010). Generating the evidence for marine fisheries policy and management. *Marine Policy, 34*, 29-35.

Hutton, G. R. G. (2009). Developing an inclusive measure of influence for marine environmental grey literature. (Unpublished Master's thesis). Dalhousie University, Canada.

Hydraulic fracturing. (2014, November 28). In *Wikipedia*. Retrieved 3 December 2014 from http://en.wikipedia.org/wiki/Hydraulic_fracturing

Khabsa, M., & Giles, C. L. (2014).The number of scholarly documents on the public web. *PLoS ONE, 9*(5), e93949.

Lawrence, A., Houghton, J., Thomas, J., & Weldon, P. (2014, November 17). *Where is the evidence? Realising the value of grey literature for public policy & practice.* Swinburn Institute for Social Research, Melbourne, Australia. Retrieved from http://apo.org.au/research/where-evidence-realising-value-grey-literature-publi-policy-and-practice

Likens, G. E. (2010). The role of science in decision making: Does evidence-based science drive environmental policy? *Frontiers of Ecology and the Environment 8,* 1-9.

Liverani, M., Hawkins, B., & Parkhurst, J. O. (2013). Political and institutional influences on the use of evidence in public health policy. A systematic review. *Plos One, 8*(10), e77404.

Luzi, D. (2000). Trends and evolution in the development of grey literature: A review. *The International Journal on Grey Literature, 1*, 105-116.

MacDonald, B. H., Cordes, R. E., & Wells, P. G. (2004). Grey literature in the life of GESAMP, an international marine scientific advisory body. *Publishing Research Quarterly, 20*(1), 25-41.

Masset, E., Gaarder, M., Beynon, P., & Chapoy, C. (2013). What is the impact of a policy brief? Results of an experiment in research dissemination. *Journal of Development Effectiveness, 5*(1), 50-63.

McKimmie, T., & Szumak, J. (2002). Beyond grey literature: How grey questions can drive research. *Journal of Agriculture and Food Information, 4,* 71-79.

McLeod, P. (2014, November 26). Nova Scotia fracking report author slams ban. *The Chronicle Herald*. Retrieved from http://www.thechronicleherald.ca/novascotia/1253658-nova-scotia-fracking-report-author-slams-ban

McNie, E. (2007). Reconciling the supply of scientific information with user demands: An analysis of the problem and review of the literature. *Environmental Science & Policy, 10,* 17-38.

Munro, R. (2013, October 4). How much science is there? *Science.* Retrieved from http://www.sciencemag.org.ezproxy.library.dal.ca/site/special/scicomm/infographic.jpg

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence: How research can inform public services.* Bristol: The Policy Press.

Parsons, P. (2010). Information overload. *New Scientist, 205*(2743), 38-39.

Pielke, R. A. (2007). *The honest broker: Making sense of science in policy and politics.* New York: Cambridge University Press.

Rajabi, F. (2012). Evidence-informed health policy making: The role of the policy brief. *International Journal of Preventive Medicine, 3*(9), 596-598.

Ravindranath, N. H. (2010). IPCC: Accomplishments, controversies, and challenges. *Current Science, 99,* 26-35.

Roux, D. J., Rogers, K. H., Briggs, H. C., Ashton, P. J., & Sergeant, A. (2006). Bridging the science-management divide: Moving from unidirectional knowledge transfer to knowledge interfacing and sharing. *Ecology and Society, 11,* 4. Retrieved from http://www.ecologyandsociety.org/vol1/iss1/art/

Sarkki, S., Niemelä, J., Tinch, R., van den Hove, S., Watt, A., & Young, J. (2014). Balancing credibility, relevance, and legitimacy: A critical assessment of trade-offs in science-policy interfaces. *Science and Public Policy, 41,* 194-206.

Scheeren, W. O. (2012). *The hidden web: A sourcebook*. Santa Barbara, California: Libraries Unlimited.

Schöpfel, J., & Farace, D. J. (2010). Grey literature. In *Encyclopedia of Library and Information Science* (3rd ed.) (pp. 2029-2039). London: Taylor & Francis.

Shanley, P., & López, C. (2009). Out of the loop: Why research rarely reaches policy makers and the public and what can be done. *Biotropica, 41,* 535-544.

Søndergaard, T. F., Andersen, J., & Hjørland, B. (2003). Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model. *Journal of Documentation, 59*(3), 278-320.

Soomai, S. S., MacDonald, B. H., & Wells, P. G. (2011a). The 2009 State of Nova Scotia's Coast Report: *An initial study of its use and influence*. Halifax: Environmental Information: Use and Influence.

Soomai, S. S., MacDonald, B. H, & Wells, P. G. (2011b). The State of the Gulf of Maine Report: *An initial study on awareness, use, and influence of the theme papers*. Halifax: Environmental Information: Use and Influence.

Soomai, S. S., Wells, P. G., & MacDonald, B. H. (2011). Multi-stakeholder perspectives on the use and influence of "grey" scientific information in fisheries management. *Marine Policy, 35,* 50-62.

Stojanovic, T. A., Ball, I., Ballinger, R. C., Lymbery, G., & Dodds, W. (2009). The role of research networks for science-policy collaboration in coastal areas. *Marine Policy, 33,* 901-911.

Thaler, A. (2010, November 15). Shades of gray: Gray literature, peer-review, and the struggle for fisheries management. Retrieved from http://www.southernfriedscience.com/?p=8641

Thelwall, M., Klitkou, A., Verbeek, A., Stuart, D., & Vincent, C. (2010). Policy-relevant webo-metrics for individual scientific fields. *Journal of the American Society for Information Science and Technology, 61*, 1464-1475.

Van der Veer Martens, B., & Goodrum, A. A. (2005). The diffusion of theories: A functional approach. *Journal of the American Society for Information Science and Technology, 57*, 330-341.

Van Noorden, R. (2014, May 7). Global scientific output doubles every nine years. *Nature.* Retrieved from http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html

Webster, J. G., & Collins, J. (2005). *Fisheries information in developing countries. Support to the implementation of the 1995 FAO Code of Conduct for responsible fisheries*. Rome: Food and Agriculture Organization of the United Nations.

Wells, P. G. (2003). State of the marine environment reports – a need to evaluate their role in marine environmental protection and conservation. *Marine Pollution Bulletin, 46,* 1219-1223.

Wells, P. G., Duce, R. A., & Huber, M. E. (2002). Caring for the sea – accomplishments, activities, and future of the United Nations GESAMP (the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). *Ocean and Coastal Management, 45*, 77-89.

Wheeler, D., et al. (2014). *Report of the Nova Scotia Independent Panel on Hydraulic Fracturing*. Retrieved from http://energy.novascotia.ca/sites/default/files/Report%20of%20the%20Nova%20Scotia%20Independent%20Panel%20on%20Hydraulic%20Fracturing.pdf

Young, S. P. (Ed.). (2014). *Evidence-based-policy-making in Canada: A multidisciplinary look at how evidence and knowledge shape Canadian public policy*. Don Mills: Oxford University Press.

Is the Licensing of Grey Literature Using the Full Palette of "Contractual" Colors?: A Comparative Analysis of Grey Literature Terms of Use

Tomas A. Lipinski,
School of Information Studies, the iSchool at the University of Wisconsin Milwaukee, United States

Andrea J. Copeland,
School of Informatics and Computing, Indiana University, United States

Introduction

This paper reviews a sample of licenses or terms of use governing the use of grey collections. A license is a form of legal permission. It need not be a contract but often is. Some terms of service reviewed do not even provide permission per se but state a list or rules users should follow. From this analysis a list of best licensing practices or checklist of considerations can be offered. Regardless of the specific character the phrase "terms of use" is used to represent the "agreements" under discussion. The framework for comparison and assessment draws upon that used in similar studies used to evaluate the terms of service for various social media sites (Lipinski & Copeland, 2013; Lipinski, forthcoming). The various terms of use are reviewed in detail and deconstructed in terms of legal essence or effect in order to present a "landscape" or descriptive overview of the legal aspects of grey literature terms of use in current use.

Research Question

What provisions are found in licenses (or terms of use) governing the use of grey literature and are these provisions adequate to preserve, make accessible and use grey literature? This will assist in fulfilling one goal of the Pisa Declaration on Policy Development for Grey Literature Resources, to "further strides in licensing grey content for both commercial and non-commercial purposes." (Declaration #10, May 14, 2014).

Research Method And Framework

Several terms of use governing Grey Literature (GL) collections are reviewed. The terms of use identified for consideration were sourced from the affiliations of GL Conference presenters since 2009 (2010-2013 or GL12-GL15) or of a grey literature collection that was a focus of a particular GL annual conference presentation. This initial review resulted in ten readily available terms of use OpenGrey (a Creative Commons based license), CNRS the National Center for Scientific Research, CERN the European Organization for Nuclear Research, Alberta Health Services (Canada), Research Gate, KISTI the Korea Institute of Science and Technology Information, STFC (Science and Technology Facilities Council, United Kingdom), British Library, Library of Zeeland (Netherlands), and the National Documentation Center of the National Hellenic Research Foundation (also a Creative Commons based license). In addition, two commercial suppliers (EBSCO Green File and Wolters Kluwer OVID) were included for comparison.

While not all terms of use reviewed fell within United States jurisdiction, the contract and other legal principles applied and discussed within are common to all countries. Second, specific terms and conditions were compared in order to identify problematic provisions in light of one possible measure of efficacy; proposing that the following minimum characteristics should be found or addressed in the required Terms of Service (TOS) in public repositories: Functionality, Integrity, Provenance and Permanence (FIPP).

As discussed in Lipinski and Copeland (2013), one framework used to assess the terms of use governing online collections consists of the following minimum characteristics: Functionality, Integrity, Provenance and Permanence (FIPP). An overview of these characteristics follows. It is based on that earlier work but is modified as indicated below to the present context, where a there is increased focus upon content availability rather than service functionality and one in which users have reduced rights to contribute content. (The prior work focused on analysis of several social network and community web services.) In this way the licensing of grey literature collections mirrors traditional licensing in libraries, i.e., database content licensing. This assessment can assist the licensors and users of licensed grey literature in designing more effective terms of use.

Functionality

Functionality, there should be…

- The ability to make a variety of uses of the available content; use rights should be at least as broad as those under the copyright law, i.e., the set of "rules" that would exist regarding the content in the absence of a license.
- Restrictions on use should be narrow and not go beyond those restrictions found in the copyright law, i.e., the set of "rules" that would exist regarding the content in the absence of a license. Where the website does allow for user contributions restrictions on content and format restrictions or exclusions should be limited at most to unlawful content, unlawful in terms of the criminal law, child pornography for example, or actionable in terms of the civil law, defamatory content for example.
- Warranty (a legal promise) given by the provider of service availability and functioning; and the potential for a warranty regarding the lawfulness of content from users.
- Limited termination provisions, including adequate notice prior to termination.

In general, two licenses used the Creative Commons (Open Grey and National Documentation Centre, National Hellenic Research Foundation). However the National Documentation Centre, National Hellenic Research Foundation included a set of terms and conditions in addition; it is those additional terms and conditions that are analyzed here. While Creative Commons is appealing for many social and legal reasons, such licenses are not without issues (Lipinski, 2013).

***User Rights and Restrictions*[1]**

License should not restrict use rights to those that would exist under copyright law, in the absence of the license. Some licenses however did just that, restricting for example reverse-engineering; a practice that in order to achieve interoperability is often found to be a fair use.[2] The Ovid terms and conditions are consistent with this position ("solely in accordance with copyright law" with specific mention of "fair use"). However, from the perspective of a user it could be argued that a license should give additional rights beyond that of the copyright law. One benefit of couching terms of use in the form of legal agreements (contractual license) is to clarify the unresolved or contentious areas of copyright and in so doing increased the scope of use rights. Licenses can for example make clear that the use of content in the creation of course packs is a fair use. Case law, at least in a commercial context suggests that such collation is not a fair use.[3] While not referencing the copyright law per se CERN has very broad use rights, allowing users to "perform work, or transmit or store data consistent with the stated goals, polices and conditions of use as defined by the bodies or bodies granting you access."

The two most common user restrictions relate to commercial use (Open Grey, Ovid, Alberta Health Service, National Documentation Centre/National Hellenic Research Foundation) and some form of data-mining or automated searching (GreenFILE: EBSCO, ResearchGate, British Library). It is arguable that such activities would be prohibited as well under broad language such as that contained in the Alberta Health Service terms and conditions ("no portions of the Site or Content…may be reproduced in any form, or by any means…") as it is typical to create a set of sub files on which to run searches when data mining.

The Library of Zeeland is the only license to use a so-called savings clause, reserving all rights for the provider or licensor that are not explicitly granted: "Any rights not expressly granted herein are reserved." Such clauses are somewhat archaic, a vestige of old contract

[1] In the previous study of co-created community repositories, it was logical to include a variable related to patron contributions to the repository. While several of the websites (National Hellenic Research Foundation and Library of Zeeland) included the right for user contributions this was not typical in the terms of use reviewed. As a result two specific variables (Ability to contribute content: create personal repository and "Content prohibited") were replaced with "Rights" and "Restrictions" respectively, within the broad category of Functionality.

[2] See, *Evolution, Inc. v. SunTrust Bank*, 342 F. Supp. 2d 943 (D. Kan. 2004) (de-compilation of software in order to extract out unprotected elements fair use); *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000), cert. denied 531 U.S. 871 (2000) (copying computer code to extract unprotected elements fair use); *Atari Games Corp. v. Nintendo of America, Inc.*, 975 F.2d 832 (Fed Cir. 1992) (reverse engineering of computer program as intermediate step to extracting unprotected elements can be a fair use); and *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992) (de-compilation of software program to extract unprotected elements fair use).

[3] See, *Princeton University Press v. Michigan Document Services*, 99 F. 3d 1381 (6th Cir. 1996), cert. denied 529 U.S. 1156 (1997).

practices. Recent courts and commentators argue that a saving clause is without legal effect, especially where use rights concern content protected by copyright. "The reservation of rights provision states: 'All rights not specifically granted to licensee hereunder are expressly reserved by Licensor." The law in this Circuit is clear, however, that such a clause adds nothing to the substantive prohibitions in the Agreement." [4]

However, a failure to elucidate rights does not take rights away. In other words, because a license fails to grant a user fair use rights for example, does not mean the user is without these rights. The fair use right is not a right for copyright holders to grant; it is a right of users. As one commentator explained: "The inclusion of a generic savings clause, such as 'All rights not specifically granted to Licensee are expressly reserved,' has no effect on a library's rights under fair use and Section 108, and hence does not bar lending under ILL arrangements. A publisher's rights are expressly limited by the exceptions in the law, including Sections 107 [fair use] and 108, so they have no right to forbid activities that Sections 107 and 108 allow."[5]

By the same token, if a provision appears to limit fair use rights in specific, such to prohibit reverse engineering, then the user will be bound to the limitations to which to he or she agreed, assuming a valid (enforceable) contract is formed.

Legal Risk and Warranty Disclaimers

As might be expected the commercial licenses (GreenFILE from EBSCO and OVID from Wolther Kluwer) contained lengthy terms and conditions. However the British Library terms and conditions were also extensive. Those terms and conditions that did form or intend to form a contract (two did not, see discussion below) included some sort of warranty waiver or "as is" clause (Open Grey, GreenFILE, Ovid, Alberta Health Service, ResearchGate, Library of Zeeland, National Documentation Centre, National Hellenic Research Foundation). An "as is" clause, as the phrase implies, indicates that the content is provided "[i]n the existing condition without modification...[in] a sale of property 'as is' means that the property is sold in its existing condition, and use of the phrase as is relieves the seller from liability for defects in that condition.—Also termed *with all faults*."[6]

Initially used by the software industry, such clauses permeate the information contracting landscape. CERN does not use this specific language but appears to intend the same result: "Use of the Grid is at your own risk. There is no guarantee that the Grid will be available at any time or that it will suit any purpose…Although efforts are made to maintain confidentiality, no guarantees are given." Similarly, STFC "disclaims all responsibility for and accepts no liability for any errors or losses caused by any inaccuracies in such information or the consequences of any person acting or refraining from acting or otherwise relying on such information."

Other warranties disclaimed in the licenses reviewed were for merchantability, fitness for a particular purposes and non-infringement (Ovid, Alberta Health Service, ResearchGate, Library of Zeeland). "[T]he implied warranty of fitness for a particular purpose is a warranty implied by law when a seller has reason to know that a buyer wishes goods for a particular purpose and is relying on the seller's skill and judgment to furnish those goods."[7] An implied warranty of merchantability is "[a] warranty that the property is fit for the ordinary purposes for which it is used."[8] One commentator offers this explanation: "The *implied warranty of merchantability* attaches when the seller is a merchant with respect to the goods involved in the exchange. Accordingly, the product must meet certain standards; it must pass without objection in the trade under the contract description and it must be fit for the ordinary purposes for which such goods are used. The concepts of marketability, operability, and repairability have emerged as varying criteria for merchantable goods."[9] Such disclaimers may appear ill-suited to information "products" such as grey literature collections or for information "services" such as access to report literatures, statistical data, etc., again reflecting the impact of software licensing practices.

[4] *Hard Rock Café, International. (USA) Inc. v. Morton*, 1999 WL 717995, *23 (S.D.N.Y., 1999) (unpublished).
[5] Brandon Butler, Appendix A: Legal Licensing Issues (REPORT OF THE TASK FORCE ON INTERNATIONAL INTERLIBRARY LOAN AND DOCUMENT DELIVERY PRACTICES (June, 2011), available at http://publications.arl.org/1acgvq.pdf.
[6] BLACK'S LAW DICTIONARY 502 (9th ed. 2009) (no pagination in Westlaw) (emphasis original)
[7] *Martinez v. Metabolife Intern., Inc.*, 6 Cal. Rptr. 3d 494, 500 (2003).
[8] BLACK'S LAW DICTIONARY (9th ed. 2009) (no pagination in Westlaw).
[9] 1 JULIAN B. MCDONNELL & ELIZABETH J. COLEMAN, COMMERCIAL AND CONSUMER WARRANTIES ¶ 1.02[1], at 1-7 (1991) (emphasis original).

A disclaimer of non-infringement means that the content supplied by the licensor does not infringe another's copyright. Under the laws of copyright liability using (printing, downloading, reposting, distributing, etc.) a work that is infringing is an act of infringement, making a user that prints, downloads, reposts, etc. liable as well. Users should be aware of such risk-shifting mechanisms in license terms and conditions. Library, educational or similar institutions should never agree to a license that does warrant (a legal promise) non-infringement, as the use of any content has the potential to trigger copyright liability, not only on the part of the library, educational entity, etc. but upon its constituents as well such as patrons, students, etc.

From the user perspective it could be argued that a warranty of non-infringement should be provided by anyone providing copyrightable content; from the provider perspective it may be unreasonable (due to cost, for example) to include such warranty when the content is provided gratuitously as is the case with many grey collections. One should never make a legal promise (warrant) that one is unsure of fulfilling. Any fear on the part of providers of grey collections built from user contributions providers should be assuaged by the various take-down procedures that exist in many countries regarding content posted by third parties that is claimed to be infringing. See, 17 U.S.C. § 512 (United States). As long as the provider expeditiously removes or disables access to the alleged infringing content posted by a third party (including a link to alleged infringing content) there will be no monetary liability whatsoever.[10] Even in the absence of so-called codified safe harbors, the general principles of secondary copyright liability would protect a provider that once on notice of the infringing nature of posted or linked content acts to remove or disable access to the infringing content. Observe that the statutory trigger for removal or disabling operates where the content may not actually be infringing. As a result it is suggested that provider make some assessment of whether the content is more likely than not to be infringing and not remove or disable unless this conclusion is reached. Even if there is a chance the content may be infringing additional protection from monetary award may exist for non-profit entities under U.S. law when reproducing content when the entity possesses a reasonable belief that reproduction of the content was a fair use.[11] Nonetheless users should be aware of using content without such warranty. Providers should be aware of safe harbor and liability principles in order to assess their level of legal risk for reproducing or hosting infringing content.

The gratuitous nature of the content provided by grey literature sites, except those published by commercial vendors of course, may also work to the advantage or content provider and obviate the need for disclaimers other than non-infringement. The nature of this harm would involve some claim of harm due to errors, omissions, or other fault with the information itself, or perhaps defamation. Most claims would rest on some concept of negligence. There are four elements to a claim of negligence: duty of care (the harm must be foreseeable in order for an actor to have a duty a care to prevent it), failure of that duty, that a failure to fulfill that duty was the proximate (or legal) cause of the harm (the harm must be foreseeable and without a superseding cause), and there must be measureable harm.[12] A plaintiff would first have to prove that there was a duty of care on the part of the provider to provide the content without any errors and omissions whatsoever or without harm of any sort. Most courts would never make this requirement of information publishers and providers, commercial or otherwise, even less so with information provided for free. The public policy implications of imposing liability would be disastrous if a court would choose to impose a duty of care that the information be 100% error free. A commercial publisher facing the reality of profits and losses and knowing that it may be liable for errors, omissions and other harms would simply opt not to provider such content altogether.

Even if there was a duty of care to provide grey literature content without errors, omissions, faults or harms of any, a court would still need to conclude that the error, omission, etc. was the legal cause of the measureable harm.[13] Where content is provided for free there can be no liability for economic harms, e.g., loss of property. Physical harms to a person may result in

[10] See, e.g., 17 U.S.C. §512(c) and (d) ("a service provider shall not be liable for monetary relief…").
[11] (See, 17 U.S.C. § 504(c)(2).
[12] The Restatement (Second) of Torts § 435, defines foreseeability as follows: "The actor's conduct may be held not to be a legal cause of harm to another where after the event and looking back from the harm to the actor's negligent conduct, it appears to the court highly extraordinary that it should have brought about the harm."
[13] "Just as foreseeability is central to finding that a duty is owed, it is also 'the touchstone of proximate cause' and of the superseding cause doctrine." *Sanders v. Acclaim Entertainment, Inc.*, 188 F.Supp. 2d 1264, 1276 (D. Colo. 2002).

liability but the gratuitous nature of the information can impact the reasonableness of the reliance upon it. The gratuitous nature of most grey literature collections means that is not reasonable for a person to rely upon it for purposes of establishing provider liability.

Is it nonetheless recommended practice to use an "as is" notice? While most legal counsel might recommend the safest course of over-protection (notices and disclaimers) the legal reality is even where the content is provided by a commercial provider, a warning notice of errors, omissions, etc. is not required to secure protection. In the seminal case of *Winter v. G.P. Putnam's Sons*, 938 F.2d 1033 (9th Cir. 1991) involving harm from publishing "*The Encyclopedia of Mushrooms*" that misidentified poisonous mushrooms as edible a U.S. appellate court concluded that a publisher has "no duty to investigate the accuracy of the contents of the books it publishes.[14] Should a provider use a warning notice to dissuade users from reasonable reliance or to investigate the accuracy of content it provide in order to meet a lesser standard of care? The court in the *Winter* answered in the negative: "with respect to the first, a publisher would not know what warnings, if any, were required without engaging in a detailed analysis of the factual contents of the book. This would force the publisher to do exactly what we have said he has no duty to do-that is, independently investigate the accuracy of the text. With respect to the second, such a warning is unnecessary given that no publisher has a duty as a guarantor."[15] Providers of information content, even commercial publishers should rest assured that courts would seldom find legal liability for errors or omissions in the grey content provided in online or other collections.

For harms other than copyright and contributed (posted) by third parties in an online context, further protections are available, at least in the United States. Federal law provides complete immunity: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." 47 U.S.C. § 230(c)(1). Section 230 provides broad immunity, even where the service provider or website would edit the content users contribute or otherwise post.[16] Contributions to grey collections by users made with the knowledge that other researchers will be able to access the content submitted should extend the immunity of section 230 to the hosting entity. As a result, the mechanisms of tort liability work in favor of publishers and providers of grey literature to protect them from liability for content the publisher or provider creates; for content created and submitted or posted by a user or other third party, as long as the content is contributed in the context of further online access by others, section 230 immunity should apply as well.

A final risk-shifting technique is to disclaim or limit the damages available should there be a dispute between the parties over harm from the product or service. A number of terms of use adopted this strategy (Open Grey, GreenFILE, Ovid, Alberta Health Service, ResearchGate, KISTI, British Library, Library of Zeeland, National Documentation Centre, National Hellenic Research Foundation).

Termination

The terms and conditions reviewed also gave the provider the right to terminate a user from its service, often upon breach (Open Grey, Alberta Health Network). The Library of Zeeland has a termination right but does not indicate what will trigger the exercise of that right ("reserves the right in its sole discretion, to terminate your access"). Recall that Ovid refused to give a warranty of non-infringement but reserves the right to "terminate the account or access of users who infringe the intellectual property rights of others." Ovid clearly desires it both ways. Both a provider and user might desire that any termination right be preceded by some notice of fault (typically a material breach) and an opportunity to remedy the fault with some time frame, say

[14] A publisher may [] assume such a burden, but there is nothing inherent in… the surrounding legal doctrines to suggest that such a duty should be imposed on publishers." Id. at 1037.

[15] Id. at 1037-1038 (underscore original).

[16] See, *Donato v. Moldow*, 865 A.3d 711, 725 (N.J.Super. 2005): "Appellants claim that Moldow controlled the content of the discussion forum, thus shaping it, as a result of which he was transformed into an information content provider. He accomplished this, according to appellants, by selectively choosing which messages to delete and which to leave posted." The statute does limit immunity to information "provided by another information content provider." In *Batzel v. Smith*, 333 F.3d 1018 (9th Cir. 2003), involving the posting on an open website of a private email that the sender did not intend to be posted the court indicated that section 230 would not apply: "immunity applies only with regard to the third-party information provided for use on the Internet or another interactive computer service." Id. at 1033.

thirty days. This notice and opportunity is known as a cure right or right-to-cure. From both the content provider (licensor) and user (licensee) perspective it is preferable to condition the right to terminate upon breach only of a material term and then only if there has been adequate notice and a right to cure, typically requiring a 30 day period in which the party in breach can cure the fault and avoid termination. See Appendix A. Functionality Provisions in Web-based Grey Literature Collections, Examples.

Integrity

Integrity (of legal particulars), requires an…

- Opportunity to view terms coupled with a clear mechanism of assent to those terms (contract formation).
- Indemnification to accompany the service and user warranties provided.
- If the licensor has the ability to change the terms and conditions of use, there should also be adequate notice of any change in terms.
- Assent (agreement) to changes should be subject to a clear and distinct mechanism, e.g., a click-to-agree screen prompt.

Most of the terms and conditions reviewed, as is typical in online settings, equate use of the service or access to content with agreement to terms and conditions of use. While this is an efficient mechanism in online settings, users should be aware of the dangers of equating use with assent; use of service by patrons might constitute assent to renewal or acceptance of a change in terms. The problem should be obvious, how is one to know when a use really means "I agree" or a use is just that, use of the service or website. Use equal assent mechanisms are prevalent in the terms of service reviewed (Open Grey,GreeenFile, Ovid, Alberta Health Services, ResearchGate, Library of Zeeland, National Documentation Centre, National Hellenic Research Foundation).

Assent to Terms and Conditions and Changes and Terms and Conditions

As observed many licenses equated use with assent, with the user bound to abide by the terms of use. Two licenses did not have its terms and conditions in the framework of a contract, i.e., the website did not intend its terms and conditions to be binding on users. In fact, the language made it clear to the user that a contract was not intended. The Science and Technology Facilities Council website statement consisting of seven lines of text state that "the information made available … does not form part of any contract between STFC and the user." The British Library website makes a similar statement: "The usage guide for images is based on goodwill. It is not a legal contract." Oddly, the British Library website still had a robust set of terms and conditions that included a number of restrictions as discussed above.

Where a contract is formed there may be a need on the part of the provider to change the terms and conditions of use. While this is common from the licensee user perspective what is more important is whether or not such changes are communicated to the user coupled with an opportunity to accept or reject (typically by ceasing use of or access to the service) the new terms. How is change of terms communicated to user of grey literature collections? Ovid uses a website posting to announce changes in terms, with further use after posting deemed acceptance of the changes. Alberta Health Services uses a similar mechanism and admonished users to "please review the Terms and Conditions regularly as your continued use of the site will constitute your agreement to any changes." The National Hellenic Research Foundation can also "amend the Terms & Conditions [] from time to time, also using a web posting mechanism. At least if the changes are "substantial" the National Hellenic Research Foundation will notify users "by posting a prominent announcement on the Website." The Library of Zeeland "reserves the right to changes the terms, conditions, and notices" but does not indicate the mechanism for effecting the changes. Only Open Grey requires that changes to terms must be based on the mutual written consent of both parties.

It is questionable whether an approach such as "check our website for changes" is a legally valid mechanism for purposes of assent to those changes. In mass market agreement scenarios such as those for grey literature collections, i.e., those scenarios where there is no opportunity to negotiate over the terms of use, where every user gets the same terms of use, at least one United States appellate court indicated that a "check our website for changes" mechanism is

indeed not valid.[17] The best way to incorporate such changes would be for a user to be unable to proceed past the home page of the website without first clicking-to-agree that the user has read and agrees to the substantial changes. Even from the provider perspective this might be preferable as then there is clear indication that the user is aware of and accepts the terms presented and that a contract is formed, at least for those provider desiring the terms be binding on users.

Indemnification

As discussed earlier a provider of content should warrant that it has the legal right to make the content available and that the content is not infringing. This should be expressed in clear language; not conditioned with limiting phrasing such as "to the best of its knowledge." The content provider should also offer an indemnification in case the promise (warrant of non-infringement) was made in error and the content is actually infringing. None of the licenses made this indemnification and as discussed above most disclaimed a warranty of non-infringement. Perhaps it is the non-commercial (free) environment of grey literature collections that is prompting this strategy, a trade-off in the minds of providers in return for the no-cost access. Even the two commercial vendors (GreenFILE EBSCO and Ovid) do not offer non-infringement warranties, Ovid specifically disclaims it. In fact, both commercial vendors not only fail to offer a promise of non-infringement and a supporting indemnification; rather both require the user to indemnify the vendor! Indemnifications for copyright harms from use of content should extend from provider to user not the other way around. Users should not enter into an indemnification lightly as it shifts along with damage disclaimers all risk from the content provider to the content user.

Some providers of grey literature do require such indemnifications from users of its content. Alberta Health Service does not require indemnification from users but does require users to "agree that any information submitted to AHS does not infringe the intellectual property of any third party." ResearchGates does in fact require users to indemnify it for "any information that is stored on request of the respective user." While this might appear logical it is somewhat unnecessary as the Section 512 take-down process—included in its terms and conditions, see discussion below—would provide for complete damages remission for infringing content posted by users. For other (tort) harms that result from user posted content, tort immunity provisions such as section 230, in the United States, provide protection as well; only the user who posted the content would be liable. Even though many collections are located in other countries, the choice of law and possibly the jurisdiction governing disputes can be the Untied States (see discussion below under Provenance). Several providers not based in the U.S. nonetheless chose U.S. law and courts as the intended jurisdiction.

In the European Union consumers have rights relating to jurisdictional issues. The rules are mandatory and cannot be waived or contracted away. The Brussels Convention on Jurisdiction and the Enforcement of Judgments in Civil and Commercial Matters for example indicates that as long as the seller's activities are directed to the consumer's domicile, the consumer can sue

[17] The United States Court of Appeals for the Ninth Circuit in *Douglas v. Talk America, Inc.*, 495 F.3d 1062 (9th Cir. 2007), cert. denied 552 U.S. 1242 (2008), concluded that posting of changes followed by use of service is not enforceable: "Talk America posted the revised contract on its website but, according to Douglas, it never notified him that the contract had changed. Unaware of the new terms, Douglas continued using Talk America's services for four years" Id. at 1065. The court concluded that the new terms were not part of the agreement: "Even if Douglas had visited the website, he would have had no reason to look at the contract posted there. Parties to a contract have no obligation to check the terms on a periodic basis to learn whether they have been changed by the other side" Id. Once a contract is formed any changes represent an offer for additional terms, an offer that in theory requires separate and distinct assent. The court further commented on the problem of providing notice unless the notice also identifies which terms changed: "Nor would a party know *when* to check the website for possible changes to the contract terms without being notified that the contract has been changed and how. Douglas would have had to check the contract every day for possible changes. Without notice, an examination would be fairly cumbersome, as Douglas would have had to compare every word of the posted contract with his existing contract in order to detect whether it had changed" Id. at 1066, n. 1.

there.[18] It is likely that users of grey collections would be considered consumers and so would have the rights of jurisdictional choice. Grey literature terms of use are not contracts for the sale of goods in the traditional sense. Article 13(3) of the Brussels Convention indicates that the rules apply to "any other contract for the supply of goods or a contract for the supply of services." Access to grey collections is a service supplied to users. A library entering into an agreement with a provider of grey literature is likely not a consumer. Rather the transaction would be considered B-2-B (business-to-business).[19] Recall also the two providers

(FTSC and the British Library) disclaim any contractual formation whatsoever so it is unclear whether protections such as those offered in the EU apply at all. For these reasons determining whether the jurisdictional provisions in the terms of use reviewed are enforceable. See Appendix B. Integrity of the Legal and Practical Circumstances in Web-based Grey Literature Collections, Examples.

Provenance

Provenance (legal) of the content should include…

- Copyright or other legal notices.
- Attribution or other designations.
- Identification of ownership rights for both service provider and user, if applicable.
- Use rights to other licensees.

Most terms and conditions reviewed made a statement of copyright ownership, i.e., that the supplied content was protected by copyright (Open Grey, GreenFILE EBSCO). Again, because all but two of the websites assessed did not allow for the contribution of user content the "Non-exclusive rights given to website or to other users content" of the prior study was replaced with "Boilerplate" as a variable of analysis under the legal Provenance area. Boilerplate refers to "[r]eady-made or all-purpose language that will fit in a variety of documents...Fixed or standardized contractual language that the proposing party often views as relatively nonnegotiable." [20]

Notices and Attributions

The use of copyright notice is required by some provisions of United States copyright law.[21] However, the concept of author or copyright holder attribution is not a copyright concept. The failure to include such attribution does not constitute copyright infringement.[22] However, since one tenet of grey literature collections is a commitment to making such works available at no cost it is not unreasonable to require in return the users of the content provide adequate citation information.

Open Grey, using a Creative Commons licensing scheme offers the strictest requirement regarding attribution and the moral rights of presentation or appearance while other terms of use included provisions related to notice. The British Library asks users to "please credit the source material" and to "preserve all public domain marks and creative commons licenses

[18] EC Convention on Jurisdiction and the Enforcement of Judgments in Civil and Commercial Matters, Brussels, 1978 O.J. (L. 304). The seller, however, may sue the consumer only in the consumer's country of domicile. Article 15 defines "consumer" as someone who is acting "outside his trade or profession" which is definition used by the European Legislature in all of its mandatory consumer rules. "Even when a contract for the sale of goods involves a consumer, the Rome Convention [Rome Convention on the Law Applicable to Contractual Obligations, June 19, 1980, 80/934/EEC, 1980 O.J. (L. 26)] permits the parties to choose governing law with one important qualification. In all cases, the consumer is protected by the law of his country of residence, if the mandatory laws (i.e., laws that a contract cannot override) of that country provide him with additional protections." GEORGE B. DELTA AND JEFFREY H. MATSUURA, LAW OF THE INTERNET § 3.04, Foreign Principles of Jurisdiction (2014) (no pagination in Westlaw). A consumer might be able to sue in his or her own country but the court may need to apply the seller's law. Other instruments echo these concepts. Council Directive 93/13/EEC of April 1993 on Unfair Terms in Consumer Contracts, 1993 O.J. (L 95) 29; and Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.

[19] See discussion regarding the library as merchant under United States contract law in Tomas A. Lipinski, THE LIBRARIAN'S LEGAL COMPANION FOR LICENSING INFORMATION RESOURCES AND SERVICES 57-58- (2012).

[20] (BLACK'S LAW DICTIONARY 502 (9th ed. 2009) (no pagination in Westlaw).

[21] See, e.g., 17 U.S.C. § 108(f)(1).

[22] Failure to credit does not amount to copyright infringement: *Graham v. James*, 144 F.3d 229, 236 (2d Cir. 1998); Moral Rights: 17 U.S.C. § 106A; *Museum Boutique Intercontinental, Ltd. v. Picasso*, 880 F. Supp. 153, 157 at n. 3 (S.D.N.Y., 1995); *Gilliam v. American Broadcasting Companies, Inc.*, 538 F.2d 14, 24 (2d Cir. 1976).

attached to the works" and "not to remove or alter any notices, conditions, forms of identification or dedications if required not to. Such requirements have support in the copyright law. In the United States for example it is unlawful to "intentionally remove or alter any copyright management information" (CMI) regarding the work, the right holders, attributes of the work, etc.[23] Consistent with this articulation, CNRS users "may not remove or change the names of authors or references to copyright of the publishers or any other means of identification of legal notice contained on the elements under license." Likewise Alberta health Service users must "ensure that all copyright, trade-mark and other proprietary notices are retained …" Even CERN, though not referencing the use a copyright notice indicates that users "shall respect intellectual property."

Boilerplate

Use of boilerplate provisions in licenses is an efficient mechanism for resolving selected issues in contract execution. Often these provisions relate to peripheral issues not essential to the central purpose of the contract, i.e., for grey collections that would be issues related to access and use of content. The typical boilerplate terms and conditions in contracts include: choice of law and choice of forum (whose law will govern disputes and where those disputes would be litigated), force majeure (also known as an act of god provision indicating that non-performance due to natural disaster and similar unforeseen circumstances will not constitute a breach of the contract), severability (or survivability, indicating that if a provision of the contract is found legally unenforceable by a court, the rest of the contract remains in force), non-waiver (whereby the decision by one party not to enforce a term or condition or act upon a breach of the other party does not preclude the party from enforcing the term or condition or taking action in the future), integration (the current agreement represents the complete understandings of the parties, superseding prior versions or any other information).

Ownership Right Claimed and Non-Exclusive Rights Granted

As might be expected where users can post content, the terms of service indicate that the one who posted the content does not surrender any rights through mere submission of content but retains any rights. Retention of copyright interests is articulated in the Library of Zeeland terms of use. Similarly, the CNRS indicates that other users cannot claim a copyright interest in the content. CERN indicates that users "shall respect the intellectual property" rights that govern the content. One would expect that providers that allow user contributions would claim a non-exclusive right to user the content submitted and that other users could make similar use, but such provisions were not apparent. It is also possible that one could argue an implied license was granted to the provider and to other uses in such instances. Many grey literature content providers do however indicate that the provider possesses a copyright interest in at least the attributes of the website (British Library, National Hellenic Research Foundation) or the content (Open Grey, EBSCO, Ovid, KRISTI) and in some cases both (ResearchGate). Making a claim of copyright in collection content is important as it indicates that misuse of content (in contravention of the terms of use) may lead not only to a claim based on contract law (for breach of contract) but also one based on copyright law (infringement). See Appendix C. Provenance Aspects in Web-based Grey Literature Collections, Examples.

Permanence

Permanence, there should be in place…

- Limitations on changes to service and content.
- A mechanism whereby content that a copyright holder claims is infringing can be removed from the website.
- Articulation of whether content removal can occur in accordance with so-called take-down processes as articulated in U.S. copyright law (17 U.S.C. § 512).
- Processes for restoration of removed content in conjunction with the take-down process.

[23] 17 U.S.C. § 1202(b)(1). In U.S. law the concept of CMI is broad enough to include the terms and conditions of use or symbols for example indicating the category of Creative Commons or that the work is protected by copyright (©). CMI "means any of the following information": name of, and other identifying information about, the author of a work, terms and conditions for use of the work, Identifying numbers or symbols referring to such information or links to such information, other information as the Register of Copyrights may prescribe by regulation. 17 U.S.C. § 1202(c)(2) and (6)-(8).

Changes in Content

Few terms and conditions reviewed allow for change in content. As might be expected the two commercial sites allow for changes in content (GreenFILE EBSCO and Ovid). As both vendors are aggregators of content and may in turn be licensees of the content each can make available, as such their provision of content is often dependant on agreements with others. Circumstances may change and the provider may no longer have the right to further make the content available. ReseachGate also "reserves the right to change, reduce, interrupt of discontinue the Service or parts of it any time." The language is broad enough to allow for changes in content. As with a change in terms the National Hellenic Research Foundation will post notice on its website; it will do the same for changes in content but it is the only service to offer this feature to users ("the sole relevant announcement of these on this website"). While it is unreasonable to expect that content never changes, a provider should provide notice to users of significant or substantial changes in content.

Removal and Restoration of Content

A number of terms of use reviewed reference either generally or by specific citation some mechanism of take-down or disabling of access to alleged infringing content. The statutory processes in U.S. law is outlined in 17 U.S.C. § 512. If a provider of grey literature includes content that a copyright holder claims is infringing the terms of service indicate that the service provider will follow the section 512 or other take-down and disabling requirement. Under U.S. law however the statutory safe harbor—protection against all monetary remedy—offers protection only for content posted by a user, not for content produced or acquired then made available by the grey literature provider.

It may be that grey literature providers employ this mechanism because as discussed earlier no warranty of non-infringement is provided and providers see this as a trade-off. Perhaps if a copyright holder realizes that a mechanism is provided for the removal of suspect (in terms of the copyright law) understands that content will be removed the copyright holder would be less likely to sue the grey literature provider. This take-down will occur even if the content is not infringing. The protection a service provider receives from the statute requires the content be removed or disabled "expeditiously." While the take-down statute provides for a process of restoration of content or access none of the TOS reviewed alert users to the possibility of restoration. Under the statute, a restoration request must come from the user whose content was removed or access disabled.

Oddly, several websites make reference to U.S. copyright law and its take-down processes even the content is not within the jurisdiction of the United States. While British Library does not make specific reference to section 512, four of the five statutory factors are indicated for the take down notice to be valid, UK address is provided for contact. The Choice of Law and Choice of Forum is of course England and Wales. The Library of Zeeland terms and conditions reference Section 512 and include "Notice and Procedures for Making Claims of Copyright Infringement." However, this makes sense as the Choice of Law and Choice of Forum is King County, Washington, USA. Perhaps the oddest combination is ResearchGate, reciting the five Section 512 requirements for an adequate take down notice, yet German contact and address information is provided with the choice of law and choice of forum being the Federal Republic of Germany. See Appendix D. Permanence Aspects in Web-based Grey Literature Collections, Examples.

Results

While there is diversity in approaches the terms of use that govern access and use of grey collections there is some consistency. However, not all practices may be either necessary from a legal risk perspective or in the best interests of promoting the maintenance and wide dissemination of grey collections as expressed in the Pisa Declaration. The adjustment in the FIPP evaluative factors demonstrate to both providers and users what legal risks still exist in license or terms of use articulations and more important what is within the realm of the possible. FIPP continues to be useful as model for both providers and users to of grey literature collections better understand the limitations and possibilities of maintenance, access and use of grey collections.

Recommendations (from the user perspective, but not exclusively)

Functionality

- Rights, should be granted to users at least as extensive as that given to users under the copyright law; this would allow for data mining for example, which is fair use under U.S. copyright law. Consider expanding the user rights in the areas of education (course packs), scholarly sharing, etc. Derivative uses should be allowed.
- Restrictions, avoid those provisions that might impact privacy rights (monitoring or reporting of abuse) or free speech rights (product reviews); restriction on commercial use is reasonable, whether content supplied by commercial or non-profit entity.
- Risk-shifting, content providers should include a warranty of non-infringement (and indemnification, see below under Integrity). Risk-shifting, "as is" warranty or other disclaimers may not be necessary; take-down and disabling mechanism may protect a responsive provider by limiting or reducing liability for infringing content posted by users; such processes may reduce liability in other scenarios. The principles of tort liability would work to protect the content provider for harms deriving from content given the gratuitous nature of grey literature content provided by government or non-profit entities. For harms other than copyright due to content posted by third parties, at least in the United States, 47 U.S.C. § 230(c)(2) provides complete immunity.
- Termination may be inevitable in some circumstances; however the right to terminate should be exercised only in situations of material breach and if notice is first given with a notice and 30 day right to cure. The use of a suspension mechanism should not be available to the provider.

Integrity

- Contract formation should be based upon a click-to-agree mechanism and provide an opportunity to read the terms before "clicking."
- Change of material terms should be accompanied by notice and click-to-agree similar to initial contract formation.
- Indemnification should accompany a warranty of non-infringement, or if not, include mechanism for take-down and disabling responses, include restoration process as well. Users should not be required to indemnify the provider.

Provenance

- Attribution is not an exclusive right of the copyright holder but use of copyright notices is often required by some statutory provisions; attribution and other information about the work including terms and service can constitute Copyright Management information which should not be removed.
- Moral rights would include attribution; while criticism of content should be allowed, use that is unlawfully prejudicial to the honor or reputation of the creator should be prohibited.
- Copyright "ownership": a user-creator does not surrender copyright in content submitted to a grey literature collection and assuming the contributor possess the legal right grants the entity a non-exclusive right to make the content available and grants users a non-exclusive right to use the content. Providers hold copyright in the website and aspects of its functionality (associated software, for example).
- Boilerplate used should include choice of law and forum (but should allow for user preferences, consistent with EU mandatory rules regarding consumer contracts), integration, severability/ survivability, non-waiver, force majeure.

Permanence

- Changes in content should be allowed, but if substantial in terms of quantity (items, articles, titles, etc.) or significance, change should be accompanied by prominent notice or alert.
- Removal of infringing or otherwise unlawful is of course acceptable, but providers should realize that various safe harbors are dependent upon removal or disabling of content before adjudication. Providers may want to make an assessment of whether or not removal or disabling is necessary in light of such request from a copyright holder. In circumstances of user contributed content removal of content should include a notice of removal or disabling to the contributor.
- Section 512 or other take-down process must include restoration rights and communication of same to contributor whose content is being removed or disabled.

References

Tomas A. Lipinski, The Librarian's Legal Companion for Licensing Information Resources and Service, 299-310 (2013) (Chapter 9: General Public Licenses, Open Source Agreements, and Creative Commons Agreements, pp. 299-336).

Tomas A. Lipinski and Andrea Copeland, Look before you License: The Use of Public Sharing Websites in building Patron Initiated Public Library Repositories, 42(4) PRESERVATION, DIGITAL TECHNOLOGY & CULTURE, at 174, November 2013 (pp. 174–198).

Tomas A. Lipinski (forthcoming), Click Here to Cloud: End User Issues in Cloud Computing Terms of Service Agreements, in IMCW2013 Proceedings (Yaşar Tonta editor, 2014) (Springer: Communications in Computer and Information Science Series).

Appendix A. Functionality Provisions in Web-based Grey Literature Collections, Examples.

	Rights (replaced: Ability to contribute content: create personal repository)	***Restrictions (Replaced: Content prohibited.)***	***"As is" and other warranty disclaimers relating to accuracy or availability of service, including damage disclaimer.***	***Termination.***
Creative Commons (Open Grey).	Nonexclusive license to reproduce, adapt, to make public display or public performance and to otherwise modify the work if necessary to access and use the work in "other media and formats"	**Commercial** use. Serial licensing required.	"**As-is**", including the absence of defects or errors or accuracy of the work. Damages disclaimed.	Automatic termination upon breach of terms.
GreenFILE: EBSCO.	Nontransferable and non-exclusive license granted to licensee for use by Authorized Users. Provide on-site and remote access to Authorized Users…"solely in accordance with copyright law," "fair use" allowed.	Data-Mining or Automated Searching ("Downloading all or parts…**systematic or regular** manner)"). Internal or personal.	"**As-is**", no warranty regarding the information. Damages are also disclaimed, representations or warranties including the absence of defects or errors or accuracy of the work.	Upon breach, **notice and 30-day cure** right required.
Ovid.	"used for your internal management, reference and informational purposes…limited local electronic copies of select materials"	"Redistribution, retransmission, publication, transfer or **commercial** or other exploitation," derivative, course or study packs, reproducing or distributing entirety or lengthy sequence and archiving.	"As is" other warranties disclaimed include: **merchantability**, **fitness** for a particular purposes and non-infringement. Damages disclaimed.	"in appropriate circumstances and at its discretions, terminate the account or access of users who infringe the intellectual property rights of others."
CNRS. (Humanities and social science information portal).	Display, download and store for a variety of uses.	Personal use. Data-Mining Combine or creative derivative. Download or print entire fascicles.	Not applicable.	Not applicable.
CERN (European Organization for Nuclear Research).	"perform work, or transmit or store data consistent with the stated goals, polices and conditions of use as defined by the bodies or bodies granting you access"	Anti-circumvention.	"Use of the Grid is at your **own risk**. There is no guarantee that the Grid will be available at any time are that it will suit any purpose…Although efforts are made to maintain confidentiality, no guarantees are given."	"entitled to regulate, **suspend or terminate** your access, within their domain of authority, and you shall immediately comply with their instructions"
Alberta Health Service.	"personal, non-commercial use"	"no portions of the Site or Content … may be reproduced in any form, or by any means…"	"**As is**" other warranties disclaimed include: **merchantability**, **fitness** for a particular purposes and **non-infringement**." Damages disclaimed. "Site provided for general information purposes only…does	"If you violate any of the Terms and Condition, your limited license to use the Content will automatically **terminate**. You must immediately cease

			not make any representation or warranty of any kind as to the quality, accuracy, currency or completeness of the Content on this Site."	your sue of the Site and **destroy any copies** of the Content or any portion thereof."
ResearchGate	None.	"**automated** or massive manual retrieval of other User' profile ('**data harvesting**')" prohibited and "information provided within the Service by the Provider and by other Users may only be accessed manually by a natural person using ordinary Internet devices."	"**as is**" other warranties disclaimed include: **merchantability**, **fitness** for a particular purposes and **non-infringement**." Damages disclaimed.	None.
KISTI (Korea Institute of Science and Technology Information).	None.	None.	"direct and/or indirect" damages disclaimed.	None.
STFC (Science and Technology Facilities Council, United Kingdom).	None (no contract formed, so no rights granted or restrictions imposed, see below).	None no contract formed, so no rights granted or restrictions imposed, see below).	"The information made available… is for information purposes…" "disclaims all responsibility for and accepts no liability for any **errors or losses** caused by any inaccuracies in such information or the consequences of any person acting or refraining from acting or otherwise relying on such information."	None.
British Library.	None (no contract formed, so no rights granted or restrictions imposed, see below).	"no part of the Content maybe **sold**, **resold**, licensed, transferred, copied or reproduced in or in part in any manner or on any media to any person…" Data-mining or automated searching ("systematically extract and/or re-utilize the content…not limited to the utilisation of any **data mining**, **robots**, or **similar data gathering and extraction tools**…" prohibited.	"does not warrant that the sharing of Content, including Public Domain or Creative Commons content will not infringe upon the rights of third parties." "disclaims all warranties, express or implied…as to the content, accuracy, timeliness or completeness...uninterrupted or error-free" Non-infringement: "makes no representation warranty , and expressly disclaims any liability with respect to copyright statements and disclaimers…" Damages disclaimed.	None.
Library of Zeeland.	"Any rights not expressly granted herein are reserved." "**By posting, uploading**…you are granting […permission to use your Submission…"	"damage, disable, over burden, or impair … or interfere with any other party's enjoyment." Numerous restrictions on content that can be uploaded: defamatory, infringing, obscene, upload viruses or corrupted files, conduct surveys, contests, pyramid	"**as is**" other warranties disclaimed include: **merchantability**, **fitness** for a particular purposes and non-infringement." Damages disclaimed.	"reserves the right in its sole discretion, to terminate your access…"

		schemes or chain letters…harvest emails.		
National Documentation Centre, National Hellenic Research Foundation.	"you may copy, distribute, transmit and adapt the work…" "if you provide any information…you agree to grant [] all necessary intellectual property rights to use such information…"	"do not use this work commercial purpose" "may not frame or mirror" not "cause damage or malfunction to the Website, adversely affect or jeopardize [] service provision"	"**As is**" "makes no representations or warranties with respect to this Website or its content, all warranties relating to this Website and/or its content and/or any website to which it is linked are here to the fullest extent by law excluded." "No representations or warranties are given as the accuracy or completeness of the information provided…" Damages disclaimed. "does not guarantee … uninterrupted… error…free of viruses"	"reserves the right to **suspend or terminate** your access and use of this Website at any time without notice."

Appendix B. Integrity of the Legal and Practical Circumstances in Web-based Grey Literature Collections, Examples.

	Contract Formation: Assent Issues	***Change in terms***	***Notice of Changes in Terms***	***Indemnification***
Creative Commons (Open Grey).	**Using** a work subject to a CC license equates to assent of terms. The rights granted to users in a CC license functions as consideration for purposes of contract formations.	If a CC license is used to govern access to content then changes to terms must be by **mutual written agreement** of the licensor and the user.	None required.	None required.
GreenFILE: EBSCO.	**Use** equal Assent.	None.	None.	Required by license for violations of Copyright Act of 1976.
Ovid.	"By **accessing, browsing and/or otherwise using**" equal assent.	Yes.	"any time without notice **by updating this posting**… you… should therefore **periodically visit** the Online Service and page to determine the then current terms and conditions of use…"	"You agree to defend, indemnify and hold harmless Ovid, its licensors, and any of their respective officers."
CERN (European Organization for Nuclear Research).	"By r**egistering** as a Grid user you shall be deemed to accept these conditions of use."	None.	None.	None.
Alberta Health Service.	"By **using** the Site, you, the user agree to the following terms and conditions."	"Terms and Conditions may be **revised** by AHS at **any time without notice**."	"**Please review** the Terms and Conditions **regularly** as your **continued use** of the site will constitute your agreement to any changes...date of the last revision…at top of page."	None, but users "agree that any information submitted to AHS does not infringe [] intellectual property rights … you grant AHS the right to provide access to the information..."
ResearchGate.	"legally binding between the Provider and any natural or legal person who…**uses** the Services	None.	None.	"Users indemnify the Provider form any claims raised by third parties to any

	for any purpose."			information that is stored on request of the respective user.
STFC (Science and Technology Facilities Council, UK).	None: "The information made available … **does not form** part of any **contract** between STFC and the user."	None.	None.	None.
British Library.	None: "The usage guide for images is based goodwill. It is **not a legal contract**."	None.	None.	None.
Library of Zeeland.	"Your **use**…constitutes your agreement to all such terms, conditions, and notices."	"reserves the right to change the terms, conditions, and notices"	None.	No indemnification, but "you warrant…that you will not use …for any purpose that is unlawful or prohibited…" and "you earrant and represent that you own or otherwise control all of the rights to your submission…"
National Documentation Centre, National Hellenic Research Foundation.	Use equals assent: "Visitors/User of this Website are bound by the Terms & Conditions…" "You will be deemed to have accepted such changes, **if you use** the Website after EKT has published the modified format and/or content and/or services on the pages of the website.."	"**may amend** the Terms & Conditions or Privacy Policy from time to time…"	"..by **posting changes** on the Website…any substantial changes …will notify you by posting a **prominent announcement** on the Website.	"you warrant to [] that such use will not infringe the Intellectual Property right of any third party." "You agree to indemnify and hold harmless EKT, its directors, employees, suppliers and affiliates, from against any damages or costs (including, without, without limitation, reasonable lawyers' fees) that arise directly or indirectly from your breach Terms & Conditions."

Appendix C. Provenance Aspects in Web-based Grey Literature Collections, Examples.

Website	***Attribution Requirements***	***Moral Rights: Presentation or Appearance***	***Copyright and Ownership***	***Boilerplate (Replaced: Non-exclusive rights given to website or to other users content.)***
Creative Commons (Open Grey).	Distributions, public performances and displays require inclusion of the Uniform Resource and applicable **copyright notices** (name of author, title, and source work if an adaptation) and keeping disclaimer of warranties intact.	Presentation rights prohibit distortion, mutilation, modification or "other derogatory action" that "prejudicial to the … honor or reputation" of the author.	Works subject to a Creative Commons license are protected by copyright.	Severability, Non-Waiver, Integration.
GreenFILE: EBSCO.	None.	None.	EBSCO or its licensees are the owners.	Force Majeure, Severability, Integration. Non-Assignment.
Ovid.	None.	None..	"materials…are copyrighted materials of Ovid or its licensors…"	Severability, Non-Waiver, Choice of Law and Choice of Forum (**New York**, USA).
CNRS (Humanities and social science information portal).	Users may not "remove or change the names of authors or references to copyright of the publishers or any other means of identification of **legal notice...**"	None.	Users cannot "acquire [] any right of ownership in the data or portions of data…	None.
CERN (European Organization for Nuclear Research).	None, but "shall respect intellectual property and confidentiality agreements."	None.	"You shall respect intellectual property and confidentiality agreements."	None.
Alberta Health Service.	"ensure that all copyright, trade-mark and other proprietary **notices are retained** on any copy of the Site materials in the same form and manner as the original."	None.	None.	Choice of Law and Choice of Forum (Province of **Alberta**, Canada), Integration, Severability, Non-Waiver.
ReasearchGate.	None.	None.	"Software…site design and other graphics, articles and other texts as well as the database are protected by copyright and property of the Provider."	Choice of Law and Choice of Forum (Federal Republic of Germany), Severability.
KISTI (Korea Institute of Science and Technology Information).	None.	None.	"All contents in the websites are copyrighted and protected by the Laws."	None.
STFC (Science and Technology Facilities Council, United Kingdom).	None.	None.	None.	None.

British Library.	"please credit the source material" "preserve all public domain marks and creative commons licenses attached to the works" "agree not to remove or alter any **notices**, conditions, forms of identification, or dedications, if required not to."	"any adaptations made to an image should not be attributed to the original creator and should not be derogatory to the originating cultures or communities."	"'the Site' … is either protected by third-party rights such as copyright or trademarks…is copyright to The British Library Board, or are materials which are in the public domain or made available under a Creative Commons License."	Choice of Law and Choice of Forum (**England** and **Wales**).
Library of Zeeland.	"Falsify or delete any author attributions, **legal or other proper notices** or proprietary designations or labels of the origin or source of software or other material…uploaded."	None.	"does not claim ownership of the materials you provide"	Choice of Law and Choice of Forum (King County, **Washington**, USA), Severability, Integration.
National Documentation Centre, National Hellenic Research Foundation.	"and you give credit to EKT"	None.	"is the copyright holder of this website"	Severability, Non-Waiver, Choice of Law and Choice of Forum (**Greek/Greece**).
	Attribution Requirements	Moral Rights: Presentation or Appearance	Copyright and Ownership	Boilerplate (Replaced: Non-exclusive rights given to website or to other users content.)

Appendix D. Permanence Aspects in Web-based Grey Literature Collections, Examples.

Website	***Changes in Content***	***Removal provision-general***	***Removal provision-specific: Section 512 take-down process***	***Restoration***
Creative Commons (Open Grey).	None.	None.	None.	None.
GreenFILE: EBSCO.	"sites may be added or deleted from this Agreement as mutually agreed upon EBSCO and Licensee.	Yes.	Yes.	None.
Ovid.	"may, at any time, remove…"	Yes.	Yes. Specific reference to Section 512 and 5 statutory requirements for a valid take-down notice.	None.
CNRS (Humanities and social science information portal).	None	None.	None.	None.
CERN (European Organization for Nuclear Research).	None.	None.	None.	None.
Alberta Health Service.	None.	None.	None.	None.
ResearchGate.	"Provider reserves the right to change, reduce, interrupt or discontinue the Service or parts of it any time.	Yes ("act expeditiously to remove or disable access t such information" "where necessary, remove or	Yes. Even though the specific Section 512 requirements for an adequate take down notice are recited, German contact	Yes. Even though the specific Section 512 requirements for a counter notice are recited, German contact information

		disable access to this information ('notice and take down' process").	information and address is provided.	and address is provided.
KISTI (Korea Institute of Science and Technology Information).	None.	None.	None.	None.
STFC (Science and Technology Facilities Council, United Kindom).	"Changing circumstances may cause STFC to have to change the information and contents of its pages at any time."	None.	None.	None.
British Library.	None.	Yes: "in the event that any Content infringes your rights or the rights of any third parties… not properly identified or acknowledged we would like to hear from you…" "material on our website, for which you have not given permission… please contact us…"	No, but 4 of the 5 statutory factors are indicated for the take down notice to be valid, UK address is provided for contact.	No. But implies restoration is possible: "material will be temporarily removed from the British Library website pending an agreed solution…following possible outcomes": replaced unchanged, replaced with changes, permanently removed.
Library of Zeeland.	"right to review material posted …and to remove any material in its sole discretion"	Yes.	Section 512 and "Notice and Procedures for Making Claims of Copyright Infringement."	
National Documentation Centre, National Hellenic Research Foundation.	"may change the format and/or content and/or services of the Website at any time and without notice, with the sole relevant announcement of these on this Website."	None, but "EKT may restrict the availability of the Website temporarily or permanently … in order to update content or for any other reason."	None.	None.
	Changes in Content	Removal provision-general	Removal provision-specific: Section 512 take-down process	Restoration

Publishing Geodesy, Topography and Cartography Research via Invenio

Jiří Drozda and **Veronika Synková,**
Research Institute of Geodesy, Topography and Cartography;
Petra Pejšová, National Library of Technology, Czech Republic

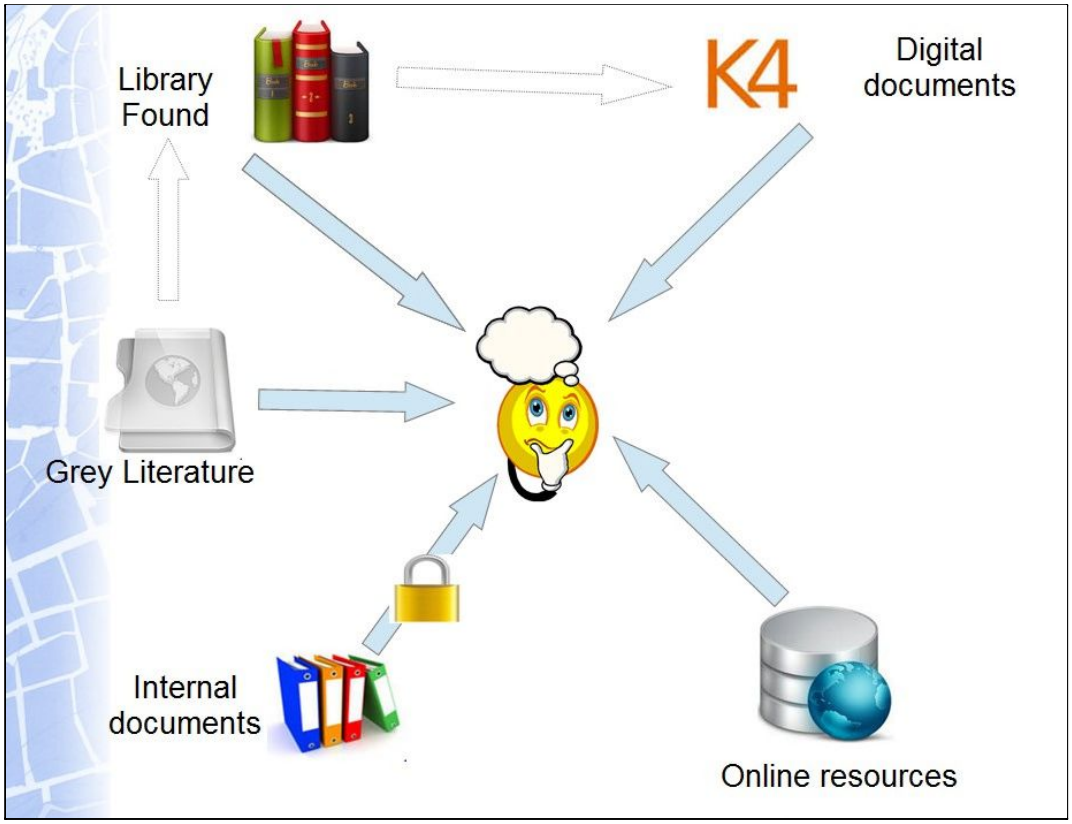
Abstract
We would like to present open source system Invenio that helps to really small specialized library to provide high standard services and grey literature open the access for limited budget and minimal staffing. It will describe access to research outputs in branch of geodesy, topography and cartography. The Surveying Library is unique library focused on geographic branch in the Czech Republic and surrounding countries. The Surveying Library keeps more than 44 thousands units/books, some of them more than 400 years old. Although the Czech Republic is a small country the community of geodesy, topography and cartography is not concentrated in one place. There are four research, ten administrative workplaces and seven universities with geographic or geodetic department. In addition the Surveying Library is located in small city Zdiby outside of Prague and can lure more travellers than students or researchers who should be main users of library. The Surveying Library found the way how to take advantage of the time when readers do not use their legs but mainly their fingers on the keyboard to reach documents they are looking for. Information technology, open source software and the common availability of the internet allowed that the library can provide their services remotely and at higher quality standards than before. In the past the information sources were fragmented. Users had to know how to reach library catalogue, the digital library, electronic resources, archives; many documents were hidden only on the webpages especially grey literature (conference proceedings, research reports, etc.). Today all information sources are provided from one common location and are part of one system in institutional digital repository. It is managed through the implementation of open source software Invenio, whose functions will be presented on the example of implementation in our Surveying Library.

Keywords: *Invenio, open source software, repository, grey literature, digital library, electronic resources, on-line library services, geodesy, topography, cartography*

Before we go into a description of the solutions used in the Surveying Library, will want to present the structure of the research institutes in the Czech Republic.
The strongest and most widely supported is a group consists of 61 research institutes of Academy of Sciences of the Czech Republic (ASCR). These institutes are financed from the budget of the ASCR and are methodically and software supported by the library of the ASCR. The library of the ASCR ensures for all research institutes ASCR collection, processing, storage and dissemination of all publications and other information sources using in R&D through the ASEP system of ASEP (Evidence of publishing ASCR activities). ASEP is a heavy library information system with a wider range of services that currently allows storing full texts. The second group is approximately 50 private research institutions financed by the companies for which they are working. This group have no interest and in many cases they cannot publish R&D outputs due to these outputs are industrial property. Last but not leastis a group of public research institutions. Those institutions are partly financed from public funds, but the main financing they must ensure through grants and projects. There is lack of methodological and technical support and it brings large problems in research management and cooperation.
Another important aspect is software used as library information systems and digital repositories. There are used Aleph, KPWin and Clavius as commercial library information systems in the Czech Republic. But those systems are expensive and system maintenance demanding for small institutions. Non-commercial library information systems are used very rarely, (Koha and Evergreen), but those system usually do not have sufficient support and are mostly “home made” system tailored for specific needs. In the last time open source information systems begin to gradually penetrate in small and medium size libraries.
The Surveying Library is part of the Research Institute of Geodesy, Topography and Cartography (VÚGTK). The library has a unique and exclusive position not only in the Czech Republic, but also on an international scale thanks to its collections and its specialization in the areas of geography,

geodesy, geodynamics, cadasters and metrology. The library is involved in many international activities involving cooperation and the provision of scientific information resources from the areas of its competence. Its uniqueness has been confirmed by the registration of "The Surveying Library®" as a trade mark (hereafter simply referred to as the library) in the Register of Trademarks held by the Czech Industrial Property Office. The library cooperates with the Czech Land Survey and Cadastral Office (hereafter simply referred to as the ČÚZK) in the collection of expert travel reports for all the employees from the ČÚZK, research reports, industry regulations and standards, all resort directives, professional newsletters and journals and the proceedings of scientific conferences and meetings.

The demise of the library was threatened a few years ago. Fortunately, the library and its unique collection managed to survive. The library has had to solve the question of how to most efficiently transform itself so that it can offer the best possible services to all its users. The main problems were in obsolete library system, fragmented information sources posted on the Web site VÚGTK. Second unhandiness was that using of most information services needs to visit the library that is located outside the city and outside of campus.



Picture 1 – Fragmented resources

The main task of library transformation was implementing of full online service. Two main objectives have been identified. Create and/or store electronic content and to give users a single access point for all services. Limitations were both insufficient founds and small number of employees. In addition, the long-standing employees had only a limited ability to learn new information systems and there was limitation in his willing to the innovation and helping in the implementation of the new system.

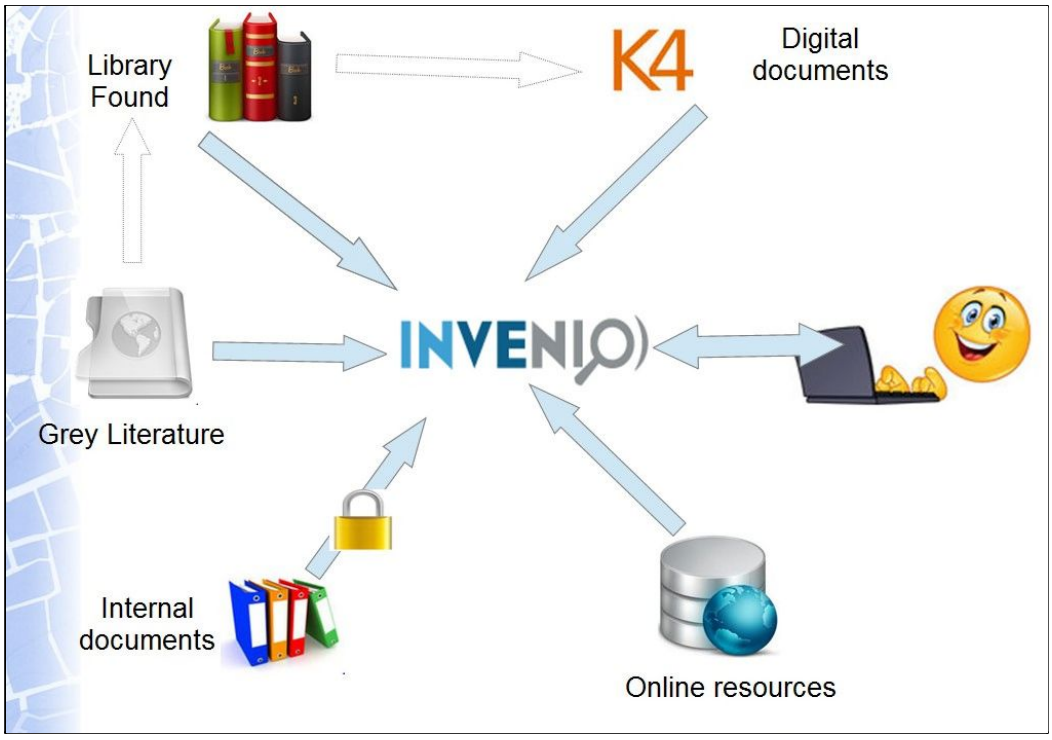
Just at the time the preparation of transformation, the VÚGTK was approached by the National Technical Library (the NTK) with the request for cooperation with the National Repository of Grey Literature (hereafter simply referred to as the NUŠL). It was a major impetus for change. The VÚGTK took up the offer NTK, which distributes the INVENIO in order to support cooperation with the NUŠL. However, the most important reason for this selection is the fact that the system library already had a built-in INVENIO module that was able to provide all the library's services and its low demand on technical, financial and personal resources.

The NTK offered a free system for processing preset INVENIO grey literature as one of the forms of cooperation. (Pejšová, 2011) The pre-setup INVENIO library system can be downloaded from the NUŠL website and installed on your own hardware. Installation instructions are also available. Furthermore, the library's INVENIO system can be edited according to the library's needs. The NTK also ensured the localization of INVENIO in Czech. Another important reason involved the fact that stable development was ensured. The European Organization for Nuclear Research (CERN) in Geneva ensures that the INVENIO system has a team of developers which develops the system, regularly publishes new INVENIO versions and has a clearly defined system development for INVENIO for the future.

The library, therefore, had to deal with two issues; it needed both a library system and a digital repository within the framework of insufficient finances and capacity building. From the perspective of maintaining the research needs of the VÚGTK, it needed to ensure the availability of a high-quality library and information resources for all its employees. The VÚGTK also wanted to act openly and to share its research output.

As well as the library's collection and the materials on the website, the VÚGTK still collects many other materials in digital form which are only available locally on the institute's network or on its employees' computers. This mainly involves the different standards of the ČÚZKC and the Surveying Office, materials from conferences, analytical and methodological materials, research data, promotional and training materials, output from projects, departmental directives, etc.

The original intention to use the INVENIO system as a library system and a digital repository has grown to an idea to use Invenio system as the discovery system too (use it as a central search point for both the printed and digital collections).

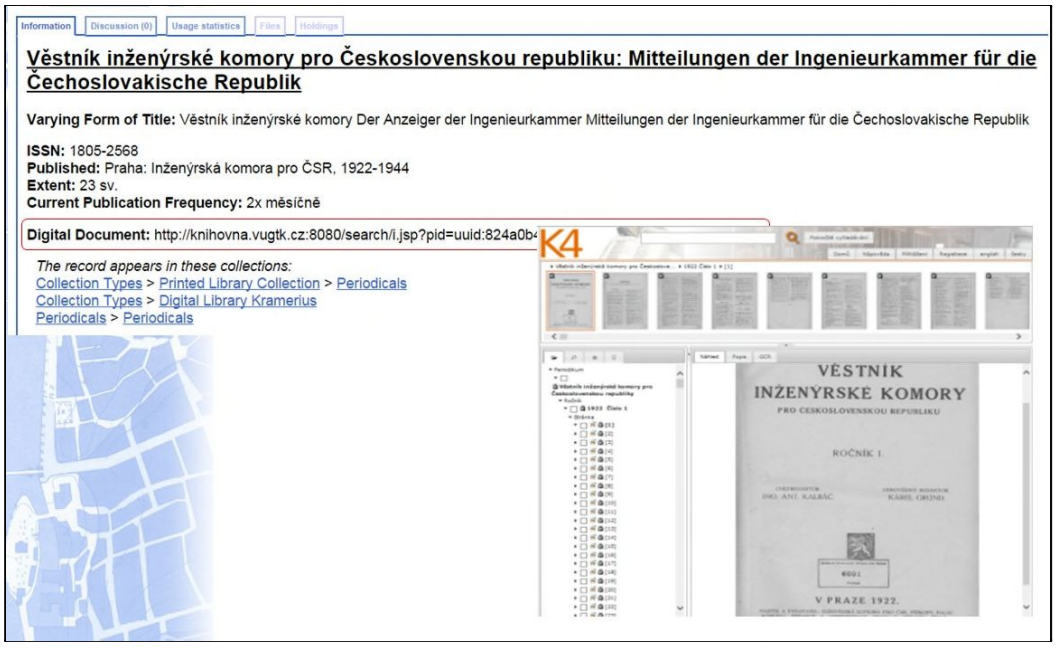


Picture 2 - The resources of the library in the INVENIO system

As already mentioned, both hardcopy and electronic documents are registered in the Invenio system. It presents all bibliographic records migrated from the original library system - 40 000 records of monographs, 3 000 periodicals records and 148 000 records of scientific articles. As a new feature, records of online magazines and digital born documents begun recorded in system. In addition recorded digital copy of printed monographs and journals are referred to a link to the digital library Kramerius.

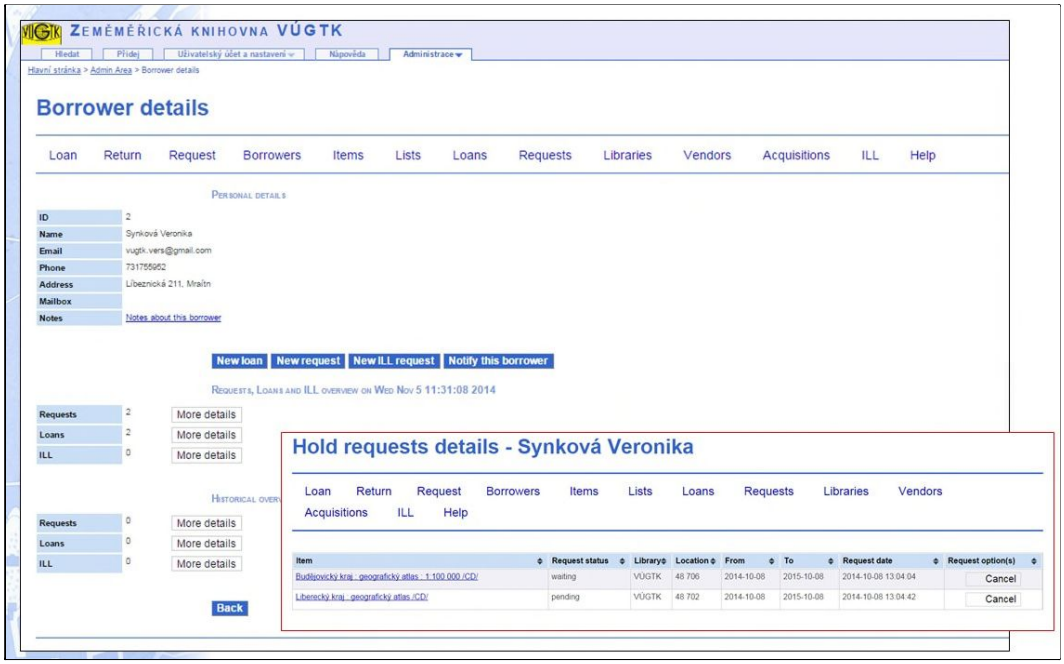
Picture 3 - Central browse and search point for both the printed and digital collections

In the case that the printed monograph or journal have been already digitised, the record is directly linked to digital Kramerius library where digital document is stored.



Picture 4 - Record with a reference to the digital library – Kramerius

A user, who has also set up access to his or her reader account via the web interface, can check the status of his or her membership account online. If the loan period has not been exceeded, the user can extend an existing loan and/or may cancel loan requirements for any titles which he or she is no longer interested in.



Picture 13 – Readers web account

Although the library has used the INVENIO system that the NTK runs for grey literature within the NUŠL project, the library has also greatly expanded its use for many other purposes. The library stores not only the grey literature but also a variety of other types of documents, including its library collection too, into the INVENIO system. Some records refer to sites outside the INVENIO system (Kramerius, the full text of available online journals and magazines).

Compared to the INVENIO version, which is run by the NTK, the library has also added some new elements such as MARC21, different templates and also the adjustment of the indexes for searching and displaying thumbnails of bibliographic records to its installation. A major extension includes the activation of the BibCirculation module and the loan service functions in this module. There has been no experience with this module in the Czech Republic to date. The activation of the BibCirculation module has brought new changes and tasks involving familiarizing the register readers with the available functions, entering orders, borrowing and returning all related instruments for the loan system. It also coincided with the need to localize the module into the Czech language. The localization has been time consuming and has still not been fully completed.

The main aim of the library is to fully satisfy both internal and external users and to always provide accessible and transparent information about the library's collection, subscribed and freely available online resources and other information materials that the library can offer all its users. The library is trying to achieve this goal by using the INVENIO library system which represents a single interface enabling searches in all the library's information sources.

The method we use is an original solution, which has not been implemented anywhere yet. This implementation also brings some technical and management complications that we are ready to address and we are willing to share our experience with other libraries.

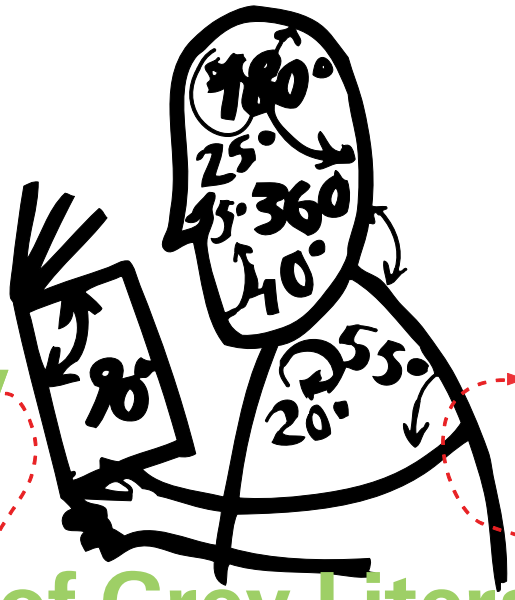
References:

PEJŠOVÁ, Petra. Národní úložiště šedé literatury. *In Seminář ke zpřístupňování šedé literatury 2011 : 4. ročník semináře zaměřeného na problematiku uchovávání a zpřístupňování šedé literatury, 25. 10. 2011*[online]. Praha, NTK, 2011. [cit. 9. října 2014]. Dostupné na WWW: <http://www.nusl.cz/ntk/nusl-97123>.

NTK

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

NUŠL



National Repository of Grey Literature in the Czech Republic

Features

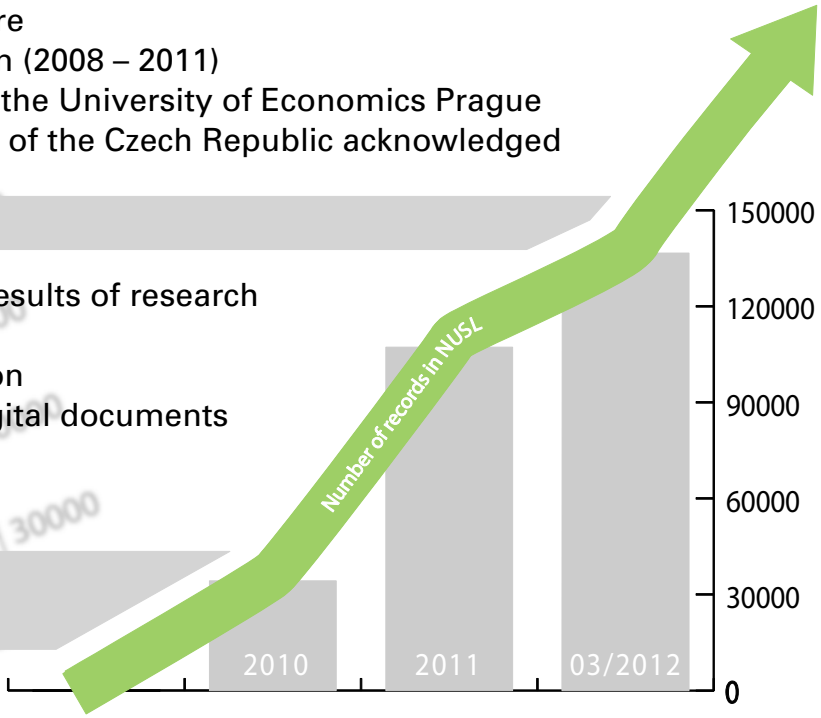
Website: www.nusl.cz
Provider: National Technical Library
Records: over 200 000 records
Partners: over 90 organizations
Source area: Academy of Science, Public Research Institutions, Universities, Libraries etc.
International Cooperation: OpenGrey, DRIVER, ROAR, OpenDOAR
Collection provenance: Czech Republic

Based on

Project: The Digital Library for Grey Literature
– Functional Model and Pilot Implementation (2008 – 2011)
Participants: the National Technical Library, the University of Economics Prague
Financial support: by the Ministry of Culture of the Czech Republic acknowledged

Goals

- Central access to grey literature and the results of research and development in the CR
- Support of science, research and education
- Systematic collection of metadata and digital documents
- Long-term archiving and preservation
- Cooperation with foreign repository



Support of expert discussion about Grey Literature

Annual Workshops:
http://nrgl.techlib.cz/index.php/Workshop
Informative Web pages: http://nrgl.techlib.cz
Publication: Grey Literature Repositories
http://nrgl.techlib.cz/index.php/Book

www.nusl.cz

Free Licences and Creative Commons: A Powerful Tool for Open Access Publishing in Grey Literature

Petra Pejšová, National Library of Technology, Czech Republic
Marcus Vaska, Knowledge Resource Service, University of Calgary, Canada

Introduction/Goal: In today's increasingly technologically savvy information society, "using remote access and free content to open doors for science students", a statement made by NANSLO lab director Daniel Branan (www.scoop.it/t/ava-openeducation), is yet another example of ongoing efforts to make information more openly and freely available and accessible. Although Branan focused his remarks on the scientific community, this applies to more than one specific subject field. Rather, scientists, teachers, artists, sociologists, programmers, as well as professionals from the arts industry and economics are increasingly becoming involved in sharing and reusing their work. Open content provides an opportunity to shorten the time for research to become available, not repeat research already conducted, have data to compare, collect background information for a project, and numerous other possibilities.

Despite the well-intentioned mandate of a Creative Commons license, the free distribution of an author's work is still "governed by applicable copyright law." (Wikipedia, n.d.) Jack Andraka, an advocate for the Open Access Movement, laments the disappointment that can occur due to publication and distribution restrictions: "I've seen so many great ideas get killed in the lab when my peers are stopped by closed access [to research articles]" (http://teamopen.cc/jack). Open licensing is a strong instrument ensuring open access to research data.

Research Method/Procedure: This project will uncover open licenses and describe how they are used, focusing on Creative Commons free licenses, the most widely known worldwide. The Open Access movement has begun gaining greater acceptance, with numerous institutions either strongly encouraging and/or requiring their faculty, students, and staff to deposit their scholarly work in the institutional repository. As a case in point, the University of Liege in Belgium established a mandate in 2008 whereby all publications must be deposited, including the full text of articles "as soon as the article is accepted by the editor" (http://www.openacessmap.org/list) . Despite the well-intentioned means of encouraging authors to deposit their works in the public domain via open-content licenses, controversy still remains that this act can alter the original author's ownership, particularly since "all transfers or licenses of copyright interests by a work's author are revocable" (Armstrong, 2010, p. 360). The University of Liege has countered this argument with their ORBi (Open Repository and Bibliography) open access repository; a clause has been added stating that access to an author's full text articles "will only be granted with the author's consent and according to the rules applicable to author's rights and copyrights" (www.openacessmap.org/list). This increased visibility in publications and access to research has resulted in ORBi currently holding a ranking of 34 out of 1746 repositories worldwide, recording more than 2 million downloads since its inception (http://orbi.ulg.ac).

Via a survey, international, national, subject, and institutional repositories will be selected, in order to determine if Creative Commons licenses are being used at these facilities and if so, how and in what way (i.e. which type of documents are being deposited?, what is the degree of usage? etc). The survey will focus on the different Creative Commons licenses available, and how these affect open access and copyright restrictions.

Results: We believe that results obtained from the survey will not only provide us with a comparative environmental scan of the existence of Creative Commons licenses at various institutions, but will also reveal insufficiencies and recommend approaches on how to increase the use of these licenses in grey literature repositories. It is anticipated that this venture will generate renewed interest and awareness in creating a more seamless link between open access publishing and grey literature. It is in this research context that the technology and innovation triangles combine, "extending the scope beyond R & D [research and development]" (Pant and Hambly-Odame, 2010), to the grey literature community as a whole. While certain document types may never be deposited into an institutional repository, and some authors may voice concerns about feeling obligated to adhere to such a mandate, the benefits clearly outweigh any potential harms. Open Access publishing in the grey literature domain via the use of Creative Commons licenses creates the multiplier effect, "permitting the creation of new works which may never have come into existence" (Armstrong, 2010, p. 368).

Open Access and the Open Access Movement: Publishing Connections

Succinctly defined, Open Access (OA) refers to material that is "free of charge, and free of most copyright and licensing restrictions" (Suber, 2013). In addition, OA documents are widely available and accessible, serving as invaluable components of research pursuits in numerous disciplines. According to Peter Suber, considered the founder of OA, the Open Access Movement began in 1993 as a result of the launch of the World Wide Web and initiation of online publishing. A decade later, several statements supporting OA began to arise, including Suber's aptly named BBB: The Budapest Open Access Initiative (February 14, 2002), the Bethesda Statement on Open Access Publishing (April 11, 2003), and the Berlin Declaration on Open Access (October 22, 2003). While the exact definition of OA differs somewhat across the statements, the underlying uniting principles remain the same: access to freely available literature without barriers, while recognizing and giving authors control "over the integrity of their work and the right to be properly acknowledged and cited" (Suber, 2013).

The continuous hunger for information, particularly that which is available at one's fingertips, culminated in the internationally renowned Open Access Week. Inaugurated in 2006, Open Access Week continues to advocate for free, immediate online access to the results of scholarly research and the right to use and re-use those results as needed (Open Access Week, 2014). Despite misgivings by some that the impact of articles published via OA is not as reputable as those found in mainstream academic journals, Suber is quick to argue that the peer review process in OA journals is just as "rigorous and honest as peer review in conventional journals, [often using] the same procedures, the same standards, and even the same people" (Suber, 2013). In order to backup these claims and offer support for publishing in OA journals, the Open Access Scholarly Publishers Association (OASPA) was founded in 2008, with a mission of representing the interests of OA journal and book publishers worldwide, in all disciplines. At present, 83 organizations and individuals are members of the OASPA including F1000Research, Utrecht University Library, BioMed Central, BMJ, Hindawi Publishing Corporation, ProQuest, SAGE, Taylor & Francis, Wiley, DOAJ, SPARC Europe, EBSCO, and many more. OASPA believes that through a shared interest in developing appropriate business models, tools, and standards to support OA publishing, "we can ensure a prosperous and sustainable future to the benefit of our members and the scholarly communities they serve" (OASPA, 2014).

Despite all the accolades for the OA Movement, particularly journals (gold OA) and repositories (green OA), Suber reminds his readers that several of the OA initiatives that exist today would not have come to fruition without publicly-funded research. Despite proclamations that OA material is free to all who wish to peruse it, Suber cautions that free is an ambiguous term, particularly since this literature is "not free to publish or produce" (Suber, 2013). To ease concerns regarding copyright along with the exorbitant fees that some vendors charge for publishing in their academic journals, organizations have been created to offset these fees, allowing authors "to publish their article in open access in a high quality journal and for a reasonable price." (Quality Open Access Market, 2014). Of particular note is the Quality Open Access Market (QOAM), an open crowd-sourcing website, marketed as a central point of contact providing authors with a wider selection of journals to publish in, and educating publishers on improving their submission and publishing policies. This creates a "transparent academic publishing environment." (Quality Open Access Market, 2014).

In a thought-provoking article published in 2013, Jeffrey Beall, a librarian at the University of Colorado, paints a different picture of what he believes are the true motives behind the OA movement. While certainly not condoning the purpose and value of OA, especially with its logical stance of freedom of information to all who seek it, Beall argues that the OA movement imposes "onerous mandates on researchers, mandates that restrict individual freedom." (Beall, 2013, p. 589). In particular, Beall singles out a number of academic journal publishers, claiming that the influx of so-called predatory journals have led to some scholarly journals losing the prestige that they once carried: "there are many unscientific ideas that people can get published in scholarly journals thanks to predatory open-access publishing" (Beall, 2013, p. 595). While Beall's views appear to be in the minority, they do posit further thought on his notion that these journals are functioning as digital repositories, leading to an increasing interest in the connections between OA publishing and the grey literature.

OA Publishing and Grey Literature

Despite an awareness among libraries and scholars of the importance of the green OA, namely institutional repositories, especially with regards to publishing in the field of grey literature, there are still only a few "institutions involved in managing repositories" (Simeonov and Stanchev, n.d., p. 165). Further, a new protocol for metadata harvesting, doajArticle, spreads awareness of the OA Movement, increasing the "interoperability between DOAJ journals and institutional repositories" (Simeonov and Stanchev, n.d., p. 167). While more and more grey literature is being made available, we have only scratched the surface of the wealth of unknown material that exists; thus access to grey literature continues and will forever remain a challenge (Banks, 2004). In a commentary published in the Journal of the Medical Libraries Association in 2004, Marcus Banks discusses the resistance that many libraries, particularly those in the health sciences, face from commercial publishers when choosing open access platforms, placing research findings in the public domain free of charge: "moral logic argues that such information should be freely available; market logic has turned it into a valuable commodity" (Banks, 2004, p. 164). Banks laments that this logic is counter-intuitive, particularly since non-published studies often have their origins in the grey literature, acting as a supplement to the core research published in the mainstream journals. Interestingly, this supplementary material is, by and large, openly accessible.

Several arguments and convictions can be made for publishing grey literature in OA domains (both gold and green). Nevertheless, despite subjecting grey literature to internal quality assessments (where a publishing institution's name and reputation are often at stake), the debate on the quality of grey literature continues, mainly due to concerns that it does not undergo the rigorous peer-reviewed process of many journals. Further, with no obligation for long-term archiving in place, grey literature continues to be difficult to locate, as it may be transferred elsewhere or become forever lost in cyberspace. Finally, "grey literature provides an essential complement to peer-reviewed findings" (Myska and Savelka, 2012); while there will always be copyright exceptions in place, free use Creative Commons licenses bypass many copyright restrictions, including being able to disseminate, re-use, or build upon an existing document. Without open access publishing, locating the grey literature would be a daunting task.

Creative Commons and Free Licenses

…"We come from a tradition of 'free culture' – not 'free' as in 'free beer' (to borrow a phrase from the founder of the free software movement[1]), but 'free' as in 'free speech,' 'free markets', 'free trade', 'free enterprise', 'free will', and 'free elections.' A free culture supports and protects creators and innovators. It does this directly by granting intellectual property rights. But it does so indirectly by limiting the reach of those rights, to guarantee that follow-up on creators and innovators remains *as free as possible* from the control of the past. A free culture is not a culture without property, just as a free market is not a market in which everything is free. The opposite of a free culture is a 'permission culture' – a culture in which creators get to create only with the permission of the powerful or of creators from the past." (Lessig, 2004)

[1]See Peter Drahos with John Braithwaite, *Information Feudalism: Who Owns the Knowledge Economy?* (New York: The New Press, 2003), 37.

In the preface to his monograph *Free Culture*, Lawrence Lessig equates freedom of information with the right to produce material openly without stringent copyright requirements or other barriers to access. Since the launch of the information superhighway, the Internet, more than two decades ago, a divide has existed between authors and their works. While most would agree with Lessig's philosophy of information sharing, concerns abound with works being re-used without giving credit or obtaining permission from the creator, particularly if the said work is being used for commercial purposes. In 2001, Creative Commons (CC) was established in Massachusetts as a "nonprofit organization that enables the sharing and use of creativity and knowledge through free legal tools." (CreativeCommons.org, n.d.). Now headquartered at Stanford University, CC has standardized the copyright dilemma by developing a set of seven licenses allowing authors to determine how and in which way their works can be shared and used. Since the inauguration of the CC movement, numerous collaborative projects in a wide range of disciplines have been undertaken, none more notable perhaps than Luke Surl's Team Open, a venture aimed at "collecting and sharing stories of the power of Creative Commons licenses" (Surl, n.d.)

Released on December 16, 2002, the set of seven CC licenses, which continue to be followed to this day, provide rights and freedoms beyond stringent copyright law and its fair use policy (Lessing, 2004). Lessing further states that in the first six months of the availability of CC licenses, more than one million objects were registered (p. 285). These seven main licenses, permitting free distribution of an otherwise copyrighted work, are as follows:

- CC0 No Rights Reserved, providing an opportunity to opt out of copyright and database protection;
- CC BY Attribution, allowing others to distribute or further build upon the creator's work, even for commercial purposes, as long as credit for the original work is given;
- CC BY SA Attribution or ShareAlike, allowing others to build upon the creator's work, even for commercial purposes, as long as the original work is credited and the new creation is licensed under the same terms as the original;
- CC BY ND Attribution – No Derivatives, allowing for the redistribution, either commercially or non-commercially, as long as the work is not changed, and credit to the original is given;
- CC BY NC – Non Commercial, allowing others to build upon the original work, for non-commercial purposes. The new work must acknowledge the creator, although as it is non-commercial, the derivative work does not need to be licensed on the same terms;
- CC BY NC SA Attribution Non-Commercial ShareAlike, allowing others to build upon the original work, for non-commercial purposes, as long as the original is credited, and the license of the new creation is identical to the original; and finally
- CC BY NC ND Attribution Non-Commercial No Derivatives. Considered the most restrictive license in the set, others may only download and share original work, as long as credit is given, but cannot change the original in any way or use it for commercial purposes.

Creative Commons & Institutional Responses Survey

Creative Commons boldly proclaims that their mission is to "help share knowledge and creativity with the world" (CreativeCommons.org, n.d.). Taking this assumption into account, the authors of this paper launched a web-based survey, with a goal of uncovering the use of open licenses, specifically CC free licenses, in digital repositories worldwide. Carefully selecting international, national, subject, and institutional repositories, the purpose of this evaluation was to determine if CC licenses were being used at these facilities, and if so, how and in what way (i.e. which type of documents were being deposited, the degree of usage of these documents and the repository itself, etc.). The survey also focused on the different CC licenses available, and how application of these terms affected open access and copyright restrictions.

The online survey ran for a one-month period, launching on September 26, 2014, and closing on October 27, 2014. Hosted via SurveyMonkey, a web-based platform familiar to the authors, respondents were tasked with answering seven questions. Although structured as a series of closed questions, necessitating only yes or no answers, respondents were probed to offer reasons for answering a question with a no. Further, nearly each question contained a comments section, thus melding the nature of the closed queries with leading open-ended requests; several respondents took advantage of this feature and elaborated on their replies.

The survey was distributed via e-mail to the administrators of 83 repositories around the world, as well as being announced at the electronic conference of the Creative Commons group, and posted on GreyNet's Linked-In social network discussion forum. Forty-five completed questionnaires were received, the results of which will now be discussed.

Results & Discussion

Question #1: Are Creative Commons licenses being used at your institutional repository? [Figure 1].

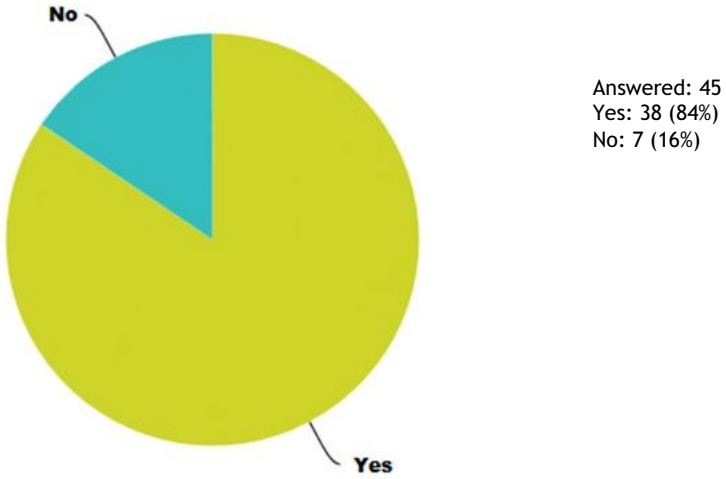


Figure 1. Level of use of Creative Commons licenses in worldwide institutional repositories.

The authors were pleased to see that CC licenses were recognized and used in the majority of institutional repositories. At sites where CC licenses were not yet established, a number of reasons were provided, the most common being legal aspects and poor awareness amongst academic staff of the existence of CC licenses and their affiliation with Open Access publishing.

Question #2: Does your institution have a Creative Commons license policy? (if yes, please briefly describe below) [Figure 2]

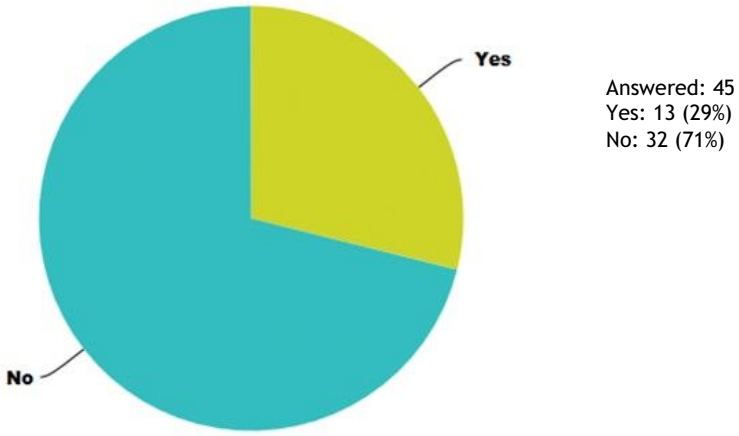


Figure 2. Availability of Creative Commons license policy in worldwide institutional repositories.

Interestingly, despite 84% of institutions surveyed claiming that CC licenses are in use within their repositories, only 29% of these organizations have established a CC license policy. An institutional CC license policy aids in explaining legal aspects while also providing authors with clear instructions on how to make best use of this policy. This creates broader awareness, requiring that each employee gain familiarity with the policy. It was thus somewhat disconcerting for the authors of this paper to learn that more than two-thirds of repositories are void of CC license standards. These numbers thus strongly support the notion that implementing a CC license policy in institutions will not only increase the use of repositories within these institutions, but will also propagate good practice in the field and thus increase awareness.

While perusing the wealth of comments provided by respondents to this question, the University of Cape Town provided an excerpt from section 9.2 of their Intellectual Property Policy, which formally endorses CC licensing:

"UCT supports the publication of materials under Creative Commons licenses to promote the sharing of knowledge and the creation of Open Education Resources. UCT undertakes certain research projects that seek to publish the research output in terms of a Creative Commons license. 9.2.1 Author(s) of Copyright protected materials that are listed in clauses 8.2 and 8.3 are free to distribute their material under a Creative Commons license. 9.2.2 Author(s) of Copyright materials that are listed in clause 8.1 should seek permission from RCIPS, who on behalf of UCT, may grant permission for the material to be distributed under a Creative Commons license."

From the remaining commentaries relayed to this question, the authors learned that policies surrounding use of CC licenses differ according to types of documents, indicating conditions of subsidy rules. Strict usage of CC licenses for all work deposited in a repository caused problems with licensing rules as agreed to with publishers, particularly in determining which CC license should be allocated to published works on a compulsory basis. However, if the publication in question was already published elsewhere, under a different type of CC license, before being deposited in the institutional repository, the original license must be honoured. A single work cannot be entered under different CC licenses.

Question #3: Under which terms is a Creative Commons License deed issued at your institution? (select all that apply) [Figure 3]

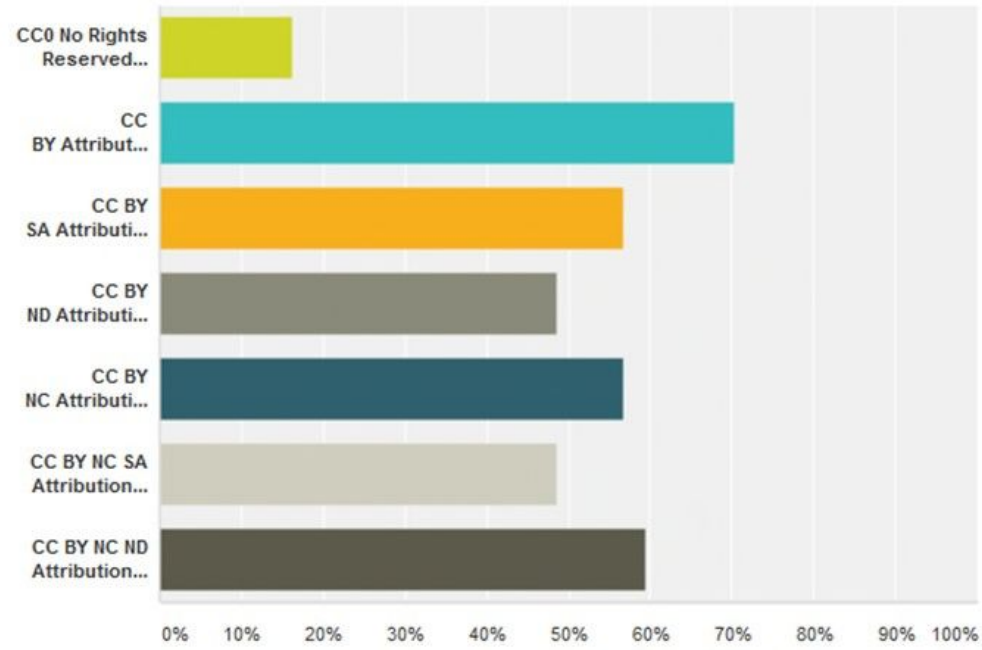


Figure 3. Terms of Creative Commons License deed in worldwide institutional repositories.

As indicated by the responses received to this question, the CC BY license is liberal, and thus used most frequently. This echoes Myska's sentiments that "the basic and most permissive is the CC-BY license. This allows all forms of distribution, copying, adapting…for commercial gain. The author however must always be mentioned." (Myska, 2013). Further, CC BY supports development of a work due in large part to freedom of usage. As Myska explains, "the licensor may also [use] restrictive license elements. Thus he may prohibit the commercial use (NC – NonCommercial), modifications (ND – NonDerivatives), or allow modifications upon condition (SA-ShareAlike)" (Myska, 2013). These types of licenses are also often used, particularly NC, NonCommercial. It can be difficult to determine what is considered commercial usage, which is one of the reasons why NC attribution has become a subject of legal arguments, despite the definition of NonCommercial use remaining unchanged.

NonCommercial use is thus understood as not primarily intended for or directed towards commercial advantage or monetary compensation. This disputed restrictive condition of the CC

licenses has been recently at least partially clarified by the interpretational guidelines published by the CC. Although not binding, they do attempt to provide at least a basic orientation for what should be considered noncommercial use. Most importantly, it is not the nature of the *subject* using CC 4.0, but the nature of *such use* that shall be decisive. Therefore, even commercial entities may use the works licensed for noncommercial uses only. Further, the NC clause does not limit the scope and exceptions provided by the respective applicable law. In addition, the licensor is not limited to use the work commercially (i.e. to make use of dual-licensing). Unfortunately, the courts rendering decisions related to this condition do not necessarily fully comprehend the proper functioning of this clause. For example, in the Curry v Audax case, the Amsterdam District Court did not award any damages to Mr. Curry. Interestingly, the court stated that CC licensed photos had no commercial value. In another CC related case, Deutschland radio, the German District Court in Cologne deemed the NC clause as not specific enough and explained it in accordance with the "Zweckubertragungslehre as 'only for private use.' However, this decision is not final as it has already been appealed." (Myska, 2014).

Question #4: What types of documents are being deposited at your institution via use of a Creative Commons license? (select all that apply) [Figure 4]

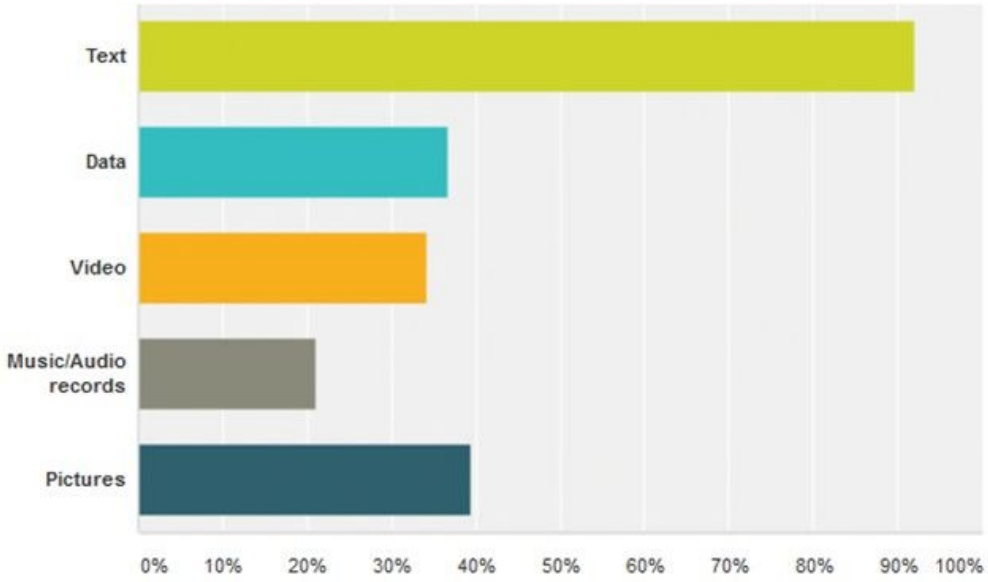


Figure 4. Types of documents deposited via use of Creative Commons license in worldwide institutional repositories.

Despite the push for increasing awareness publishing in different mediums, texts continue to occupy the majority of document types (92%) deposited in the repositories of the institutions surveyed. This finding seems to run parallel to various forms of grey literature: despite a noted increase in visual and audio representation, theses, government documents, conference proceedings, etc., all in text report format, continue to prevail. Nevertheless, despite the majority of text documents, pictorial and video elements are on the rise, perhaps attributable to increase trends in the use of social media.

Question #5: How often are Creative Commons licenses used at your institution?

In correlation with the first two questions in the survey, the authors of this paper were curious to see if any parallels could be drawn between awareness of CC licenses, subsequent policies to encourage their use, and frequency of use. Despite 84% of respondents claiming use of CC licenses in their repositories, it appears that only 16% of these licenses are used on a regular basis. More than one-half (54%) of replies indicate that CC licenses appear to be used on an ad hoc basis, causing concerns in consistency and application.

One of the respondents to this question fittingly mentioned the need to respect the rights of third parties: "almost always and whenever possible…sometimes a CC license cannot be applied where the document in question contains third party copyright material." As the results of the authors' survey prove, there is considerable worldwide awareness and attention to CC

licenses, and yet there continues to be a divide between claims that CC licenses are easy to use and practical compared to notions that the provision of training materials will create greater awareness.

Question #6: Besides Creative Commons licenses, does your repository use any other freely available licenses?

The majority of respondents (70%) indicated that CC licenses were exclusively used within their institutions. However, for those organizations that perused other freely available licenses, the range of possibilities was fast, and included General Public Licenses (GNU/GPL.AGPL/Free Documentation License), Open Database Licenses (ODbL), Public Domain Mark, UK Open Government License, Free ART License; Non-Exclusive Distribution License and a Metadata Open License.

Question #7: Please share your experiences using Creative Commons Licenses

The final question on the survey was qualitative in nature, open-ended to allow free reign and personal comments for understanding what CC licenses meant to readers. More than half of the respondents (25) provided thought-provoking replies. Some institutions are currently on the cusp of initiating a repository, others have had repositories in place for a number of years, and some continue to be wary of the purpose of CC licenses, expressing fears that they do not wish for others to modify any aspects of their works. Nevertheless, the majority of voiced opinions supported the CC movement, which the authors of this paper view as a positive trend towards open access publishing in the grey literature realm. Perhaps then it is fitting to conclude with the comments of one user and his/her experience with DSpace, a common institutional repository platform:

"In general, authors/submitters do not pay much attention to the Creative Commons license screen in our DSpace workflow and simply by-pass reading the description. We have had some instances where authors submit papers that have contradictory copyright statements on their title page and we have to contact them individually to explain the terms of the CC license and to request that the statement be removed or changed to reflect the terms to which they agreed in the default license. To date, no author has objected to the terms of the CC license. We make a point of emphasizing the importance of articulating re-use terms. Some awareness among faculty and students has been aided by the advocacy and education practices of our Copyright unit. However, there is some discussion that our default license is too conservative, particularly for the purposes of re-using the data, and that we will need to create more policies, documentation, and outreach around recommended CC licenses for different types of content. We considered it a great victory to be able to incorporate the CC license as a default in our repository license and I don't think it likely that we will be able to generate buy-in for a more open default license at this time. My institution has an open access statement but no mandate."

Conclusion

"Previously the domain of a few champions and committed individuals, but usage is on the increase". This comment from a survey respondent fully captures the current state of using Creative Commons licenses, where usage of such licenses in digital repositories remains high despite political underpinnings. Open access statements with no accompanying mandate are problematic, and legal aspects of CC license usage, coupled with pour knowledge of their existence and correct usage is one of the primary reasons why 84% of repositories surveyed refrain from using CC licenses on a regular basis. An environmental scan of the literature available on CC licenses and their use in repositories mirrors the sentiments of the gatekeepers who replied to the authors' survey questions; experiences have been either positive or cautious, but no comments or case studies were outright negative. This fact supports efforts to spread awareness of CC licenses not only to the grey literature community, but to all researchers worldwide. The authors of this paper thus recommend CC Licenses as a standard for publishing grey literature material. Reflecting back on the types of CC Licenses in use today, and in accordance with results from the survey of institutional repositories, CC BY Attribution is most widely used, and due to its characteristics, likely most suitable. This allows for the widest possible dissemination including commercial use which the grey literature community should not hold back from.

References

Armstrong, T.K. (2010). Shrinking the commons: Termination of copyright licenses and transfers for the benefit of the public. *Harvard Journal on Legislation, 47*(2), 359-423.

Banks, M. (2004). Connections between open access publishing and access to gray literature. *Journal of the Medical Library Association 92*(2), 164-166.

Beall, J. (2013). The Open-Access Movement is not really about open access. *TripleC 11*(2), 589-597.

Lessig, L. (2004). *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York: Penguin. Retrieved November 16, 2014 from http://www.free-culture.cc/freeculture.pdf

Myska, M. (2014). *Creative Commons 4.0*. Retrieved November 16, 2014 from http://www.nul.cz/ntk/nusl-175812

Myska, M. (2013). *Public Licenses, Grey Literature, and Research Data.* Retrieved October 8, 2014 from http://www.nusl.ca/ntk/nusl-161463

Myska, M. (2012). A model framework for publishing grey literature in open access. *JIPITEC 4,* 104-115.

Open Access Scholarly Publishers Association (OASPA). *The International Community of Open Access Publishers.* Retrieved October 8, 2014 from http://oaspa.org

Pant, L.P., and Hambly-Odame, H. (2010). Creative Commons: Non-proprietary innovation triangle in international agricultural and rural development partnerships. *The Innovation Journal: The Public Sector Innovation Journal, 15*(2). Retrieved March 5, 2014 from http://www.innovation.cc/scholarly-style/pant_odame_creative_commons4final2rev.pdf

Quality Open Access Market (QOAM). (2014). *About QOAM.* Retrieved November 20, 2014 from http://www.qoam.eu/about

Simeonov, G., and Stanchev, P. (2011). *Open Access and Institutional Repositories in Bulgaria*. Retrieved November 6, 2014 from www.math.bas.bg/infres/IS-publ/IS-20110Simeonov_Stanchev-DiPP.pdf

Suber, P. (2013). *Open Access Overview*. Retrieved October 8, 2014 from http://legacy.earlham.edu/~peters/fos/overview.htm

Suber, P. (2009). *Timeline of the Open Access Movement*. Retrieved October 8, 2014 from http://legacy.earlham.edu/~peters/fos/timeline.htm

Surl, L. (n.d.). *Team Open*. Retrieved October 30, 2014 from http://teamopen.cc/all/

Wikipedia (n.d.). *Creative Commons License*. Retrieved October 8, 2014 from http://en.Wikipedia.org/wiki/Creative_Commons_license

Slovak Centre of Scientific and Technical Information **SCSTI**

Achieve your goals with us

CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR

INFORMATION SUPPORT OF SLOVAK SCIENCE

SCIENTIFIC LIBRARY AND INFORMATION SERVICES

- technology and selected areas of natural and economic sciences
- electronic information sources and remote access
- depository library of OECD, EBRD and WIPO

SUPPORT IN MANAGEMENT AND EVALUATION OF SCIENCE

- Central Registry of Publication Activities
- Central Registry of Art Works and Performance
- Central Registry of Theses and Dissertations and Antiplagiarism system
- Central information portal for research, development and innovation - CIP RDI >>>
- Slovak Current Research Information System

SUPPORT OF TECHNOLOGY TRANSFER

- Technology Transfer Centre at SCSTI
- PATLIB centre

POPULARISATION OF SCIENCE AND TECHNOLOGY

- National Centre for Popularisation of Science and Technology in Society

IMPLEMENTATION OF PROJECTS

- National Information System Promoting Research and Development in Slovakia – Access to electronic information resources - NISPEZ
- Infrastructure for Research and Development - the Data Centre for Research and Development - DC VaV
- National Infrastructure for Supporting Technology Transfer in Slovakia - NITT SK
- Fostering Continuous Research and Technology Application - FORT
- Boosting innovation through capacity building and networking of science centres in the SEE region - SEE Science

www.cvtisr.sk
Lamačská cesta 8/A, Bratislava

Degrees of Openness: Grey Literature in Institutional Repositories

Joachim Schöpfel, Charles de Gaulle University Lille 3
Hélène Prost, CNRS, Associate member GERiiCO, France

Abstract
In spite of the growing success of the open access initiative, a significant part of scientific and technical information remains unavailable on the web or circulates with restrictions. Even in institutional repositories created to disseminate the scientific production of an academic institution, broad and open access to more or less important sectors of the scientific production is restricted. In order to provide new empirical evidence, 25 large institutional repositories from different continents were selected in the international directory OpenDOAR. For each repository, the access to the full text for different document types was evaluated, and the statistics were analysed for each site and cumulated. Building on our past work and new empirical data from large institutional repositories on different continents, we distinguish between different degrees of openness. Which are the main reasons, which are the stabilizing functions of this situation? The communication tries to provide some elements of understanding, together with good practices and recommendations.

Introduction

The institutional repository of the University of Nebraska-Lincoln[1] (UNL) is a collaborative service of the university's libraries that aims to provide long-term preservation and world-wide electronic accessibility of digital materials deposited by faculty, researchers and students associated with UNL. It contains 13,296 theses and dissertations from 1897 to 2014 (October 6, 2014). 12,024 dissertations are indicated as "available in PDF" while 1,272 dissertations have no link to full-text. Yet, for "non-UNL users" it is impossible to log into the UNL proxy server and download any of these PDF files, and they have to request the items through interlibrary loan.

The open access principle requires that scientific information be made widely and readily available to society. Defined in 2003 as a "comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community"[2], open access implies that content be openly accessible and this needs the active commitment of each and every individual producer of scientific knowledge.

In spite of the growing success of the open access initiative, a significant part of scientific and technical information remains unavailable on the web or circulates with restrictions[3]. Even in institutional repositories created to disseminate the scientific production of an academic institution, broad and open access to more or less important sectors of the scientific production is restricted.

Institutional repositories (IR) have been defined as "tools (…) for collecting, storing and disseminating scholarly outputs within and without the institution" (Jain, 2011) and as "a set of services (…) for the management and dissemination of digital materials created by the institution and its community members (based) on organisational commitment to the stewardship of these digital materials" (Lynch, 2003). With 2,258 sites out of 2,729 (83%)[4], they represent the most important part of the so-called green road to open access.

They contain many scientific documents that were not available previously on the Internet, but some items are under embargo or restricted to on campus access, and for other items there is only metadata, without links to the full text. Following the OpenDOAR, many repositories contain different document types (table 1).

[1] DigitalCommons@University of Nebraska-Lincoln, available at http@digitalcommons.unl.edu
[2] Berlin Declaration on Open Access, available at http://openaccess.mpg.de/Berlin-Declaration
[3] Estimations on the part of scientific publications in open access are difficult and vary widely, between 10% and 50%, depending on type of documents, discipline, country and source of information.
[4] Statistics from the OpenDOAR directory (28 September 2014).

Document type	% of institutional repositories
Articles	72%
Books, book chapters	37%
Electronic theses and dissertations	61%
Communications	37%
Reports	(37%)*
Working papers	(37%)*
Patents	4%
Datasets	4%

Table 1: Part of institutional repositories containing different types of documents (N=2,258) (*OpenDOAR index reports and working papers together as "unpublished")

Most of the institutional repositories contain articles (72%) and theses (61%). One third have books, conferences or unpublished papers (reports, working papers…). Very few hold patents or datasets (4%). Normally, all these items should be at least "gratis", available, ready to view, read and download, if not in "libre" open access with maximum reuse rights. In fact, one part of them are neither gratis nor "libre". Open archives are less open than they should be.

Yet, it is difficult to estimate the part of "missing" full text in institutional repositories. OpenDOAR warns that "full texts are not available for most results" of its content search tool but does not provide any statistics. Operated by the Bielefeld University Library, the search engine BASE harvests metadata from more than 50 million documents but indicates that the full text is available for only 75% of them.

These are global figures. Recently, Ahmed et al. (2014) reported low availability of electronic theses and dissertations in several Asian institutional repositories. In order to contribute to a better understanding of this situation we conducted a survey of 25 institutional repositories together with more than two million items. Our intention was to evaluate their degree of openness with specific attention to different categories of documents. Some results have already been published (Schöpfel & Prost 2014). The following paper shifts the focus on grey literature.

Methodology

The empirical data in our study are from a sample of 25 institutional repositories. All repositories were selected using OpenDOAR, the authoritative directory of academic open access repositories. The following search criteria were applied:

- Repository type: Institutional
- Content type: PhD theses and articles (at least)
- Size: 10,000+ items (preferred)

The search was conducted by region (Europe, Asia, Africa, Australasia, North America, South America/Central America/Caribbean), and only those repositories that are operational (i.e. recently updated), that contain different document types including non commercial literature (theses, reports etc.), that allow for filtering by document type and access options (full-text vs. restricted/no access to full-text) as a browse and/or search functionality and that indicate the exact number of results (retrieved items) were selected.

Secondly, we conducted a detailed search and/or browsed on each site for specific document types: articles, books and book chapters, conference proceedings and communications, reports, PhD theses, and working papers (unpublished). We also looked for patents and datasets but did not include them in the global analysis. For each document type, we distinguished the items with free and non-restricted access to the full text (open access) from those with restricted access (embargo, intranet, authorized users, etc.) or without full text (reference only). Whenever possible, we also made this distinction for the entire repository content.

The repositories were selected in February 2014. The analyses of each site were conducted in February and March 2014.

Results

Size and openness of the repositories

The selected repositories (IR) are listed in the appendix. For our study, we did not evaluate the whole content of each IR but limited the analysis to six document categories (working papers, theses, reports, articles, communications, books/book chapters). The total number of items in our study is 2,086,622.

The median size of the sample repositories is 26,683 documents, ranging from 1,199 (Amherst) to 775,561 (HAL[5]). Again, this is not the total size but the sum of the selected and evaluated document types, excluding for example courseware, images or Master dissertations. Thus, the true size of the IRs is higher (in average, +9%).

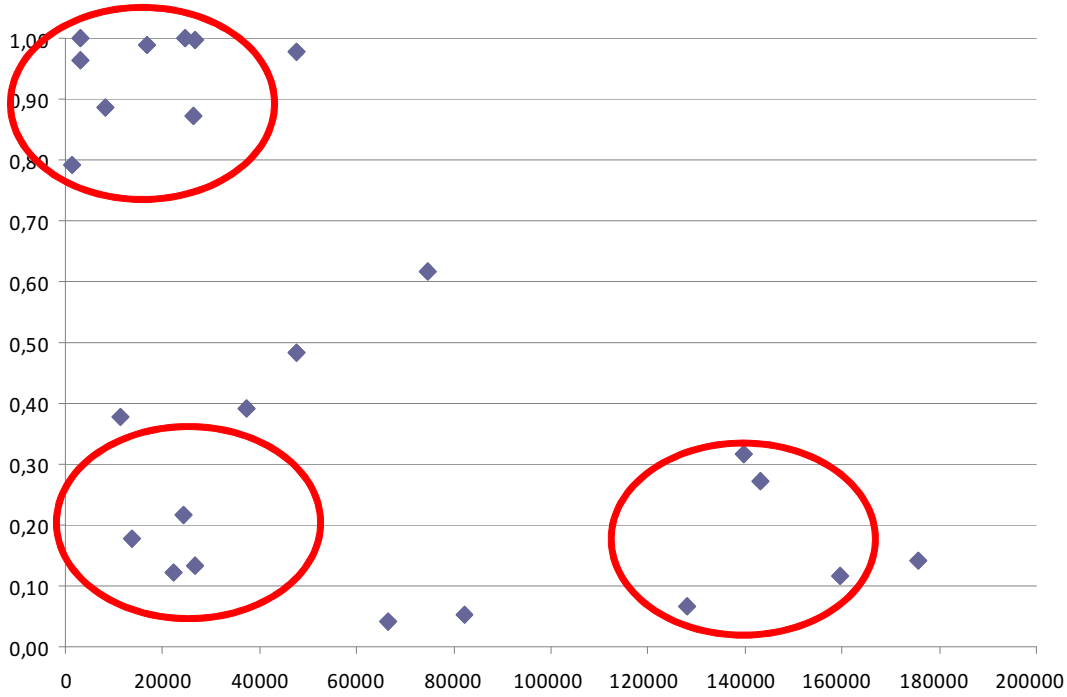


Figure 1: Openness and size of institutional repositories (without HAL)

The median degree of openness of all repositories is 0.38, which means that only close to 2/5 of all items provide open access to the full text. The individual repositories range from 0.04 (only 4% of items have full text) to nearly 1.00 (except for a few items, all deposits have freely available full text). There is no significant correlation between size and openness; yet, all larger repositories have degrees of openness below the median, while the repositories with higher degrees of openness (higher than the median) are generally smaller. Figure 1 shows three different clusters of repositories: smaller repositories with low level of openness (left side, below), smaller repositories with high degree of openness (left side, above), and larger repositories with lower degrees of openness (right side, below). HAL is part of this third cluster, with nearly 800,000 items and a degree of openness of 0.37. Why is there no large repository with a high degree of openness, that is, with a large number of metadata linking to full text? Is access restriction the price to pay for the development and growth of repositories? At least and obviously, it is a risk.

Openness per document type

The sample contains 53% of articles and 10% of books and book chapters, i.e. items that are not generally considered as grey literature. The other document types included in the analysis are distributed as follows (figure 1).

[5] The French HAL repository is a special case insofar as it is a national open repository that integrates several institutional repositories from universities and other research organisations. It was included in the study for this reason.

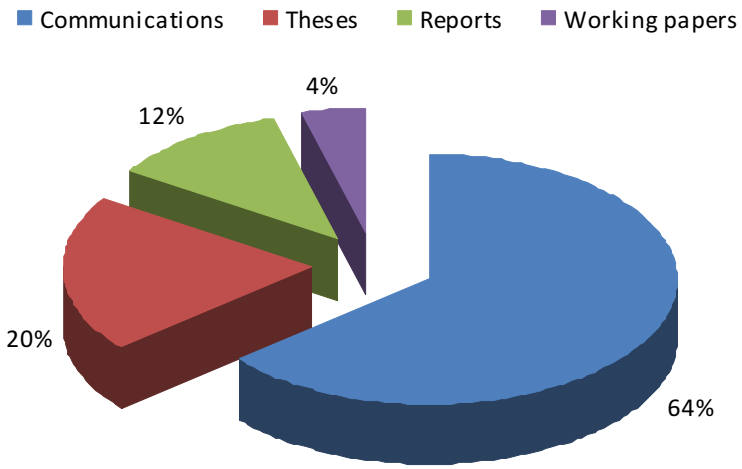


Figure 2: Document types in the institutional repositories (only grey literature, N=769,237)

In addition to these items, some repositories also contain patents and datasets. These items represent 2.8% (datasets) and 0.4% (patents) of the global content.
The evaluation of their degree of openness – the part of the items freely available on Internet – offers specific values for each document type.

Document type	Number of items	Degree of openness
Communications	490,442	0.21
Theses and dissertations	156,546	0.78
Reports	91,069	0.63
Working papers	31,180	0.96

Table 2: Degree of openness per document type, with number of items

The overall degree of openness of working papers is 0.96, which means that in the entire sample all but 4% of the working papers are freely accessible, followed by PhD theses (0.76) and reports (0.63). Significantly less open are communications (0.21) (table 2).
The median degree of openness per repository confirms the overall statistics. The median is high for working papers (0.98) and theses (0.92), medium for reports (0.63), and low for communications (0.29). The variance of openness (dispersion from average) is low for working papers, medium for theses and reports, but high for communications (figure 3).

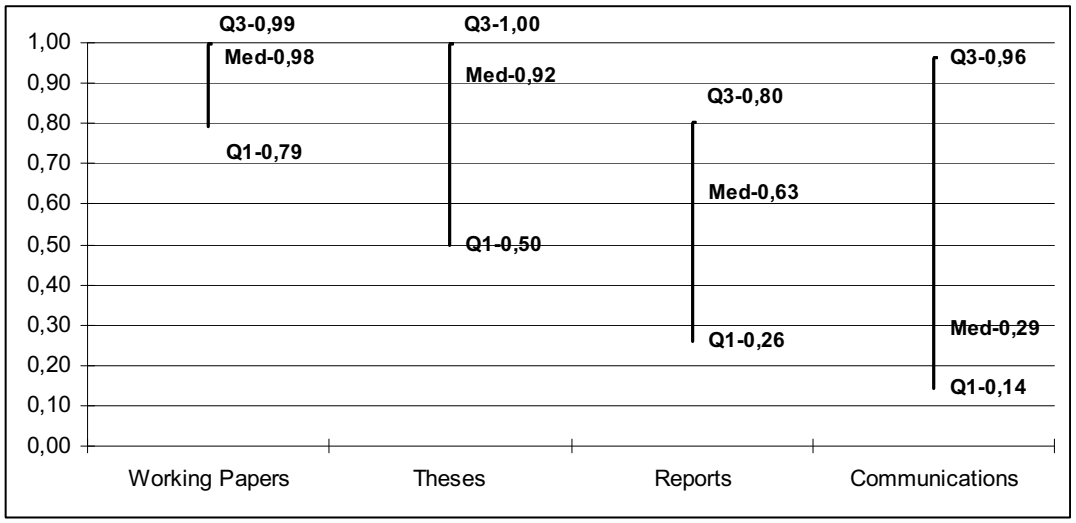


Figure 3: Degree of openness per document type with Median, 1st and 3rd quarter (only grey literature, N=769,237)

However, we must be careful with interpretation because all of the repositories have theses, most have reports and communications, but only half of them have working papers, a fact which reduces the variance.
A last observation: the number of items and their openness are inversely correlated, in that the more important categories (communication) are less open than the less important ones (table 2). Yet, this inversed relationship is not significant.

Datasets and patents
Some institutional repositories contain datasets and patents. These items are not really grey literature – datasets are not literature, and patents are not hard to find. But as they are not available through usual publishing channels, they are sometimes considered as "in the margin" of grey literature. We identified nearly 70,000 items in our sample; 60,219 datasets and 8,982 patents. While only 3% of the datasets were freely available, patents are disseminated with a degree of openness of 0.61, which means that nearly 2/3 of the patents are freely accessible in these repositories.
This is surprising for two reasons: because of the global tendency in favour of "open data", research data in institutional repositories should be more available but are not, obviously; patents are often protected behind pay walls but in our sample, several thousands are freely available.

Embargo and restricted access
Often the real nature of access restriction remains uncertain. Are the documents under embargo and will they be released and openly accessible in the future? Are they restricted to on-campus access only or is it both of these? And what about missing full text, records without documents? From our results we can only make a cautious guess, that embargo periods represent a small part of access restrictions (only 2% in our sample) and that most of the lack of openness is caused by on-campus only access and by the deposit of metadata without a corresponding document. Embargo decisions are taken in particular for PhD theses while very few reports and even less working papers are embargoed.

Repositories with a high degree of openness to grey literature
In our sample, we identified seven repositories with high levels of openness to all kinds of grey literature, i.e. with more than 80% items with full text in each category (table 3).

	Working Papers	Theses	Reports	Communications
DIGITAL.CSIC	*0,99*	*0,93*	*0,99*	*0,81*
Frankfurt a M	*1,00*	*1,00*		*1,00*
Milano	*0,99*	*0,83*		*0,97*
Dokuz Eylül University Izmir		*1,00*		*1,00*
Chiba	*1,00*	*0,99*	*1,00*	*0,96*
UNTexas		*0,92*	*1,00*	
Stellenbosch		*1,00*		*1,00*

Table 3: Institutional repositories with high degrees of openness (type 1)

Obviously, these repositories apply an open access policy that prefers unrestricted availability of documents to other objectives, in compliance with the initial goal of direct communication of the open access movement. In particular, the institutional repositories from the Spanish CSIC and Chiba University (Japan) are exemplary and should be taken for best practices, such as those from the Universities of Frankfurt and Milano even if they do not contain (or index) reports.

Other repositories
Six repositories have relatively low level of openness for all document types (table 4). Here, they clearly adopt a policy that prefers exhaustiveness (completeness) to openness. Is this still open access or do these repositories become tools designed to increase impact and facilitate evaluation of the scientific output?

	Working Papers	Theses	Reports	Communications
Torino		*0,48*		*0,04*
Ghent		*0,39*		*0,18*
INRA		*0,15*	*0,27*	*0,15*
Uppsala		*0,39*	*0,21*	*0,03*
Chalmers		*0,17*	*0,32*	*0,26*
NTUR		*0,55*	*0,28*	*0,02*

Table 4: Institutional repositories with low degrees of openness (type 2)

The other repositories are somewhere in between, with higher degrees of openness for some document types and lower degrees for others. Typically, theses and working papers are "more open" than reports and communications (table 5).

	Working Papers	Theses	Reports	Communications
HAL	*0,98*	*0,97*	*0,73*	*0,24*
KNAW	*0,79*	*0,28*	*0,65*	*0,32*
Geneva		*1,00*	*0,74*	*0,85*
Macquarie University Sidney	*0,99*	*0,50*	*1,00*	*0,20*
Swinburne	*0,34*	*0,93*	*0,26*	*0,24*
Monash	*0,54*	*0,75*	*0,02*	*0,01*
RMIT		*1,00*	*0,00*	*0,07*
Brisbane	*0,93*	*0,64*	*0,82*	*0,67*
HKU	*0,80*	*0,99*		*0,13*
SMU		*1,00*	*0,07*	*1,00*
UMassAmherst		*0,70*	*0,65*	*0,87*
Western Kentucky University		*1,00*	*0,60*	*1,00*

Table 5: Institutional repositories with mixed degrees of openness (type 3)

There are probably various explanations and reasons for these "IR profiles", combining ETD policies, disciplinary particularities (working papers) or a large number of communications published by commercial publishers. Figure 4 illustrates the average differences of these three types of repositories.

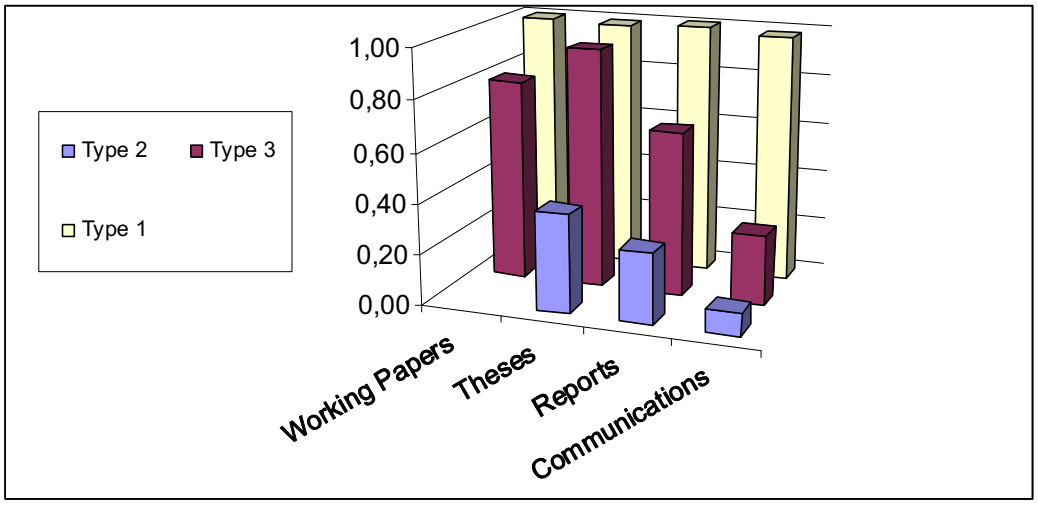


Figure 4: Three types of institutional repositories (median of openness)

However, if we compare this reality with past expectations about the Internet as "the solution for grey literature", we must admit that this is right for one part of the institutions and collections but not for all. A significant part of communications, reports and even theses still remain "hidden" and "hard to get" items.

Discussion

Methodological shortfalls

Three aspects limit the reliability of our data: (1) The different repositories do not index their content in the same way, document typologies are not consistent, and the interpretation of items as "grey literature" is not always easy. (2) Statistics on the content of repositories are often not available or not complete; thus our sample is already a somewhat biased selection of "best practice sites". (3) Only one part of the repositories clearly inform about the nature of restricted or denied access to the full text.

Typology of institutional repositories

Institutional repositories can be described and distinguished in many different ways. Our survey invites to four different typologies:

Size: Even if the focus of our methodology puts is on larger repositories, the sample includes the whole range of repositories[6]:
Small (<1,000): 1 repository
Medium (1,001-5,000): 8 repositories
Large (>5,001): 16 repositories

Openness: Following our data, we would suggest three categories:
Low degree of openness (<0.30): 11 repositories
Medium degree of openness (0.30-0.80): 6 repositories
High degree of openness (>0.80): 8 repositories

Document profiles: The openness criteria can be further differentiated by document types (theses, working papers etc.). As we showed above and limited to the field of grey literature, we can distinguish three different repository types (figure 4). Including commercial publication, this typology can be quite different and shed another light on repositories (see below).

Repository types: In the past, we suggested four different types of institutional repositories (Schöpfel et al. 2012). Even if we did not match our sample against this typology, it seems obvious that some of them are similar to type 4 repositories ("institutional deposit") while others are more like type 1 ("publishing grey literature") or type 3 ("scientific heritage") repositories.
These different typologies may allow …… in the future profiling of open repositories, in order to improve knowledge and understanding of open access but also to enhance marketing and service development.

Grey and white literature

We conducted the same analysis for published articles, books and book chapters (Prost & Schöpfel 2014). Compared to grey literature, their degree of openness is generally lower:

	Openness
Working papers	0.96
Theses	0.78
Reports	0.63
Articles	0.31
Communications	0.21
Books, book chapters	0.17

Table 6: Degree of openness per document type (N=2,086,622)

[6] For the categorization, see Kindling & Vierkant (2014)

Only one third of all articles in our sample repositories is available in full text and without any restrictions, and less than 20% of all books and book chapters can be downloaded by any user. Obviously, regarding openness there is a significant advantage of grey literature over so-called white (commercially published) literature. Figure 5 shows this advantage on the level of each repository.

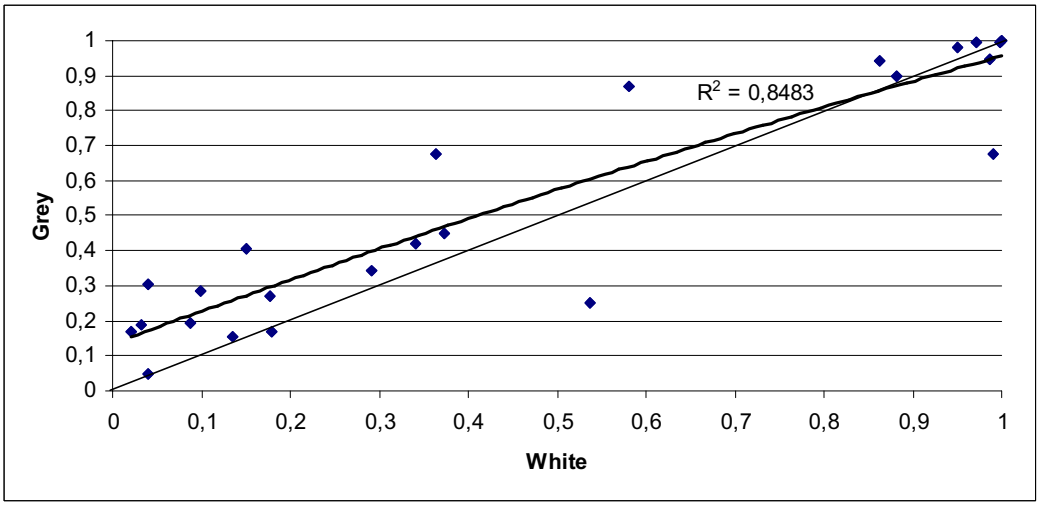


Figure 5: Degree of openness for grey (vertical) and white literature (horizontal) per repository

Except for two repositories (Chalmers and Amherst), the cumulated degree of openness of grey literature is at least as high as for published items; most often, it is significantly higher. The reason for the two "special repositories" seems different; Amherst is a small repository where most items, in particular articles, are in open access while the availability of some dissertations and reports is restricted to the campus. Chalmers appears to accept (and probably foster) deposits that cannot be disseminated freely on the Web so that especially dissertations, reports and communications cannot be accessed off campus. The tendency shows a strong relation between the two variables (determination coefficient R^2=0.85).

The special case of conferences

Yet, these figures and statistics also confirm the special case of communications. Their overall degree of openness is not only lower than for other types of grey literature but they are similar to published articles and books. Figure 6 shows the same statistics as figure 5, except for communications which have been deleted from the cumulated degree of openness for grey literature. The result is obvious.

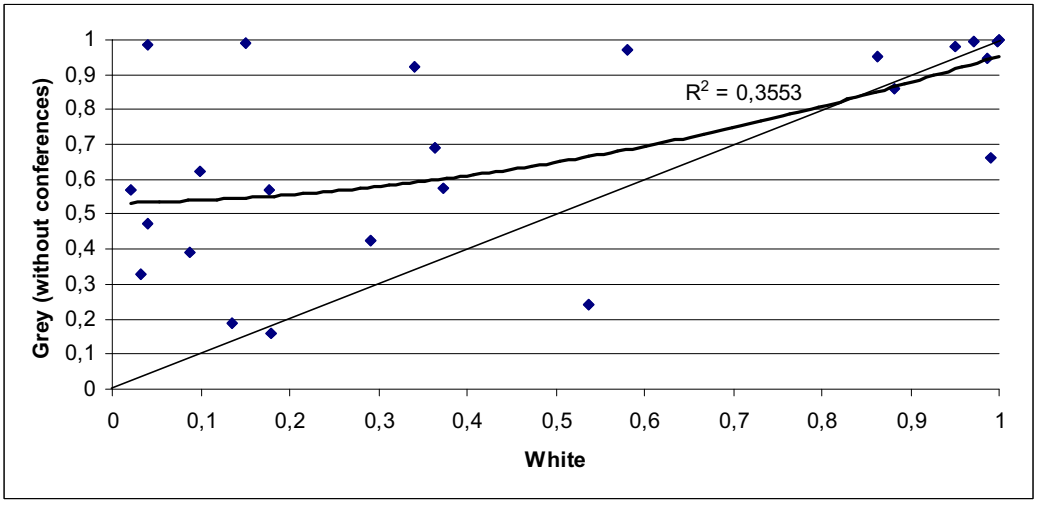


Figure 6: Degree of openness for grey (vertical) and white literature (horizontal) per repository (without communications)

This time, without communications, the difference between grey and white literature is more significant, and the tendency line with an R^2=0.36 confirms a relatively low relationship between both document types. The message of this figure is clear: institutional repositories are good for the availability and open dissemination of grey documents, at least better than for published items.

So what about communications? The reason is probably that an important part of communications is edited and disseminated by commercial publishers, in special issues of journals together with articles or as proceedings like monographs. Therefore, one part of them is not grey but white. Following our figures, this percentage can be estimated at 40-60%.

About embargo

Embargo periods and other access restrictions are normally enforced by publishers, in order to protect their revenues by delayed availability of "their" items in open archives, i.e. articles, books and those communications published by commercial publishing houses.

Our figures show that one part of grey literature is also embargoed or limited to on-campus availability, without any pressure from publishing companies. The reasons are different and have nothing to do with licensing or publishing conditions, at least not directly. Privacy and confidential information may play a role, such as third party rights, fear of plagiarism and, especially for datasets, competitive strategies opposed to free sharing of results with everybody and not only with colleagues.

Yet, publishers' open access policies indirectly affect one part of decisions on embargo and on-campus access, because PhD students anticipate these policies and often prefer not to publish in open access before a formal publication of their dissertation by a corporate publisher.

Stabilizing functions

According to published studies and our own surveys, we can distinguish between three main reasons that tend to stabilize the unsatisfying situation with restricted access to grey literature in institutional repositories:

1. For one part of these grey items, especially for communications, the authors have probably transferred their rights to publishers that disseminated the conference proceedings as special journal issues or monographs. This may explain why a significant number of communications are under embargo or cannot be disseminated outside of the campus.
2. In particular for PhD theses, authors may prefer restricted or no access at all because they intend to publish the content as a book with commercial publishing houses.
3. Last explanation, the preference of some hosting organizations for an exhaustive number of metadata that allows for evaluation and studies of the institutional scientific output. Here, the institutional repository becomes a tool for evaluation and a showcase for the institutional productivity, like a web-based bibliographic database while the need of the scientific community and society for access to results (full text, data) is not or only partially respected.

As we showed for PhD theses, embargo decisions can be motivated by many different reasons and people (Schöpfel et al. 2014). For the scientist in need of information, this is all but satisfying.

Best practice

However, these stabilizing functions are not always barriers to open access. Even in the small sample of our survey, some institutional repositories perform better than others, with higher degrees of openness. Benchmark studies should reveal their way of dealing with these problems. Why are they better, i.e. more open? Probably, they are different on six dimensions:

1. Mandatory policies (institutional support, acquisition policy).
2. Selection of deposits (moderation, metadata policy).
3. Specific approach for different document types (ETD policy, working paper publishing…).
4. Institutional workflows (including assistance for submission).
5. Legal environment.
6. Commitment to open access principles.

In the field of PhD theses, such kinds of studies will be undertaken by the European H2020 project ETD4OA.

Conclusion

In the past, we put forward the fact that for grey literature, "open is not enough", i.e., institutional repositories need a set of minimum requirements for grey items such as metadata, selection procedures, quality standards, collection management and clear deposit policy (GL13, Schöpfel et al. 2012). We then narrowed our research on electronic theses, recommended five ways to add value to theses in open archives (GL14, Schöpfel 2012) and identified access restrictions to theses in institutional repositories (GL15, Schöpfel & Prost 2014). This means that while "open is not enough", all deposited theses in repositories are nevertheless not open; or more specifically, they are available with different degrees of openness, and some are not available at all.

This year, we return to a larger perspective. Building on our past work and new empirical data from large institutional repositories on different continents, we make distinctions between various degrees of openness. Based on our empirical data, our recommendations for improved access to grey literature in institutional repositories would be:

- Typology: standardize the description of document types in institutional repositories, with a common and accepted terminology. Metadata should clearly index the document type, in terms derived from a controlled and standard terminology.
- Discovery: allow browsing and searching with document types.
- Access rights: clearly indicate the availability or access restrictions for each document, differentiating (at least) between embargo and on-campus access. Metadata should include precise information on access rights (Neylon et al. 2014).
- Policy: foster the deposit of metadata with full text for grey literature whenever possible, in particular for reports, working papers and theses. Submission of metadata without full text should be an exception, not by default. For the scientific community, access to documents is more important than exhaustiveness of records that are already published and available elsewhere on the Internet.
- Differentiation: distinguish between the different document types.
    - Reports and working papers should be systematically and by default made fully available in open access, and access restrictions should be the exception.
    - Incentives should help research teams to publish their working papers on their institutional repository and not (exclusively) on other web sites.
    - Proceedings should be published either in open access or with rights that allow immediate open access, at least after a short embargo period (six months). Institutional repositories should be able to publish proceedings as a collection and not only in the form of individual papers; perhaps they should also be able to manage the preparation and selection of conference papers (at least they should be linked to these systems).
    - A specific institutional workflow should be created for theses and dissertations in order to facilitate and foster their publishing in open access.

For grey literature in institutional repositories, unrestricted, open access should be the status by default, not the exception or an option among others. Non (or less) controlled by commercial publishers, grey literature should not be disseminated under embargo or on the campus only. Of course, embargoed items are better than no items at all, and "gratis" open access is better than no access at all. But we should not call open what is not open and transform vectors of direct scientific communication into tools of evaluation and control of institutional output.

References

A. Ahmed et al. (2014). `Theses and dissertations in institutional repositories: an Asian perspective'. *New Library World* 115(9/10).

P. Jain (2011). `New trends and future applications/directions of institutional repositories in academic institutions'. *Library Review* 60(2): 125-141.

M. Kindling & P. Vierkant (2014). `2014 Census of Open Access Repositories in Germany, Austria and Switzerland'. In *Open Repositories 2014*. Helsinki, Finland, 11 June 2014.

C. A. Lynch (2003). `Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age'. Report 226, ARL Association of Research Libraries.

Neylon, C. et al. (2014). 'Standardized metadata elements to identify access and license information'. *Information Standards Quarterly* 26(2): 35+.

H. Prost & J. Schöpfel (2014). `Degrees of Openness. Access Restrictions in Institutional Repositories'. *D-Lib Magazine* 20(7/8).

J. Schöpfel (2012). `Five Ways to Add Value to PhD Theses, and Two Ways to do it Better'. In *GL14. Fourteenth International Conference on Grey Literature: Tracking Innovation through Grey Literature*. Rome, Italy, 29-30 November 2012.

J. Schöpfel, et al. (2012). `Open is not enough. A case study on grey literature in an OAI environment'. In *GL13 Thirteenth International Conference on Grey Literature: The Grey Circuit. From Social Networking to Wealth Creation*. Washington, DC, 5-6 December 2011, pp. 75-86, Amsterdam. TextRelease.

J. Schöpfel & H. Prost (2014). `Back to Grey. Disclosure and Concealment of Electronic Theses and Dissertations'. In *GL15 Fifteenth International Conference on Grey Literature. The Grey Audit: A Field Assessment in Grey Literature*. CVTI SR, Bratislava, Slovak Republic, 2-3 December 2013, pp. 110-118, Amsterdam. TextRelease.

J. Schöpfel, et al. (2014). `Restricted vs. Open Access for Electronic Theses and Dissertations - A Challenge for Public Science'. In *ETD 2014. 17th International Symposium on Electronic Theses and Dissertations*. Leicester, July 23-25, 2014.

Appendix – List of surveyed repositories

Europe

Chalmers http://publications.lib.chalmers.se/ Chalmers Publication Library contains research publications produced at Chalmers University of Technology, Göteborg, Sweden.
CNRS http://hal.archives-ouvertes.fr HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research papers, including nearly 100 institutional repositories from French HE and research institutions.
CSIC http://digital.csic.es/ Digital. CSIC the institutional repository of the Spanish National Research Council (CSIC)
Frankfurt a. M. http://publikationen.ub.uni-frankfurt.de publication server of Goethe University Frankfurt am Main
Geneva http://archive-ouverte.unige.ch/ Open Archive UNIGE (University of Geneva)
Ghent https://biblio.ugent.be/ Ghent University Academic Bibliography
INRA http://prodinra.inra.fr ProdINRA institutional repository of the French National Agricultural Research Institute
KNAW http://depot.knaw.nl Repository of the Royal Netherlands Academy of Arts and Sciences
Milan http://air.unimi.it AIR Archivio Istituzionale della Ricerca of the University of Milan
Torino http://porto.polito.it PORTO open repository of publications produced by the scientific community of Politecnico di Torino
Uppsala http://uu.diva-portal.org institutional repository of the Uppsala University

Australasia

Macquarie University Sidney http://www.researchonline.mq.edu.au Research Online open access digital collection
Monash University Melbourne http://arrow.monash.edu.au Arrow research repository
Queensland University of Technology Brisbane http://eprints.qut.edu.au/ QUT e-Prints Archive
RMIT Royal Melbourne Institute of Technology http://researchbank.rmit.edu.au/ RMIT Research Repository
Swinburne University Melbourne http://researchbank.swinburne.edu.au Swinburne Research Bank

Asia

Dokuz Eylül University Izmir http://deu.mitosweb.com/ open archive
Chiba University http://mitizane.ll.chiba-u.jp/curator/ CURATOR Chiba University's Repository for Access To Outcomes from Research
University of Hong Kong http://hub.hku.hk/ HKU Scholars Hub institutional repository
Singapore Management University http://ink.library.smu.edu.sg/ InK Institutional Knowledge at Singapore Management University
National Taiwan University http://ntur.lib.ntu.edu.tw/ NTUR National Taiwan University Repository

America (North, Central and South America, Caribbean)

University of Massachusetts Amherst http://scholarworks.umass.edu/ ScholarWorks@UMass institutional repository Amherst
Western Kentucky University Bowling Green http://digitalcommons.wku.edu/ Top Scholar institutional repository
University of North Texas Denton http://digital.library.unt.edu/ UNT Digital Library

All websites were accessed in March and April 2014.

Korea Institute of Science and Technology Information (KISTI)

English version - http://en.kisti.re.kr/

*** Vision**

World-class information research institute creating values for customers

*** Main functions**

*** Management and service of Korean R&D reports**

KISTI exclusively manages, preserves, and serves Korean R&D reports for citizens and government officials. It provides Korean R&D reports and their information with National science & Technology Information Service (NTIS) and National Discovery for Science Leaders (NDSL).

***Contact information**

KISTI email address: hcpark@kisti.re.kr

Headquarters: Tel : +82-42-869-1004, 1234 Fax: +82-42-869-0969

Analysis of Collection and management of the Korea National R&D Report

Kiseok Choi, Cheol-Joo Chae, Yong-hee Yae, and Yong Ju Shin,
Korea Institute of Science and Technology Information, KISTI, Korea

Abstract

The term 'National R&D Program' has been widely used in a Korean society, but it is quite new as a legal term. Recently, its legal concept from the perspective of the unified semantics was defined, and the regulation on its scope has been established. In particular, 'National R&D Program' was officially used as a legal term for the first time in 'Special Act on Innovation in Science and Technology (No. 5340) enacted in April 1997. In the Regulation on the Management of National R&D Programs, 'National R&D Program' is defined in Article 2 as follows: "An R&D program in science and technology, in which a certain R&D project is chosen and partially or wholly funded by the government or public organization in accordance with the related law. It refers to all programs except for the basic programs promoted by government-funded research institutes in accordance with Paragraph 1 of Article 2 of Act on the Establishment, Operation and Fostering of Government-funded Research Institutes". In particular, because the national R&D report, one of intangible research outcomes among national R&D programs, contains important matters on R&D such as R&D method and results, it is the final outcome of the national R&D program regardless of its type. Since it is the evaluation standard of the R&D project and major research outcome as the final outcome which takes place at the end of the R&D program, the systematic management of related information is especially important. In this sense, the establishment of the information distribution system which makes it possible to collect, manage and use national R&D reports has been perceived as one of the most important R&D policies in any country. In fact, each country has systematically managed related duties through a national science & technology information agency. After all, the management of national R&D reports is a critical part in the management of the research outcomes of national R&D programs. This study examines the trend of national R&D report production, progress & assignment of national R&D programs and current non-disclosure reports, investigates the current collection and management of national R&D reports by the KISTI, a professional R&D report management agency in Korea, and proposes a future direction. In addition, the development of an information distribution system and database for national R&D reports and online services are analyzed.

1. Introduction

For a national R&D report, the result of the national R&D programs, diverse systems have been developed and operated to disclose the public knowledge asset to the general public and spread its outcome. In other words, a general distribution system on the related information has been established and operated through National Science & Technology Information System. In addition, a national R&D report is regulated to be registered on the research outcome management agency (Korea Institute of Science & Technology Information, KISTI) as a part of the national R&D outcomes to have it kept and managed by the government.

The term 'National R&D Program' has been widely used in a Korean society, but it is quite new as a legal term. Recently, its legal concept from the perspective of the unified semantics was defined, and the regulation on its scope has been established. In particular, 'National R&D Program' was officially used as a legal term for the first time in 'Special Act on Innovation in Science and Technology (No. 5340) enacted in April 1997. Even though 'National R&D Program' was used as a legal term, it just meant 'government-led R&D program'. Then, its legal concept was clearly defined in Framework Act on Science and Technology (No. 6353) enacted in January 2001. After stipulating the clauses on 'Promotion of National R&D Programs (Article 11)', 'Regulations on the Management of National R&D Programs (Presidential Decree No. 17429)' were enacted in December 2001 in accordance with the provision above. Then, the legal terminologies were finally defined and established. [1] According to Article 2 of 'Regulations on the Management of National R&D Programs', the National R&D Program is defined as follows: "An R&D program in science and technology, in which a certain R&D project is chosen and partially or wholly funded by the government or public organization in accordance with the related law. It refers to all programs except for the basic programs promoted by government-

funded research institutes in accordance with Paragraph 1 of Article 2 of Act on the Establishment, Operation and Fostering of Government-funded Research Institutes". [2] The fact that the basic programs promoted by government-funded research institutes are excluded means that the managerial characteristics of national R&D programs promoted by government-funded research institutes are admitted, not that the characteristics of national R&D programs are denied.

Therefore, the National R&D Program can be defined as follows: "A government-led R&D program which develops and promotes a plan for the advancement of national science and technology using government or public funds with a particular purpose". More specifically, it refers to "An R&D program in science and technology, in which a certain R&D project is chosen and partially or wholly funded by the government or public organization in accordance with the related law". [3]

2. The production type and management of national R&D outcomes

2.1 Production type of national R&D outcomes

The outcome of national R&D programs refers to all research results which occur after the fulfillment of the research projects regardless of their type and period. From the broad perspective, it refers to "both scientific and technological outcomes which are created through R&D such as patents and papers and other tangible and intangible economic, social and cultural results".

The outcome of national R&D programs can be largely divided into two categories: tangible outcome which is obtained during the execution of national R&D programs such as research equipment & materials, research facilities and prototype and intangible outcome including industrial property right and copyright on research reports.

First, tangible outcome refers to various results obtained from the byproducts of R&D, including biological resources, compounds and information data. Their performance is measured based on quantity or usage. On the contrary, intangible outcome is knowledge performance acquired through R&D, including knowledge assets (ex: patent, etc.), and its performance is measured considering both quantitative and qualitative aspects. In terms of royalty, the performance is analyzed based on revenue.

Meanwhile, the science and technology research outcome is divided into primary and secondary outcomes depending on its development stage or period. In other words, it is classified by the concept of time-based outcome depending on time difference. The primary outcome refers to the result which directly occurs for a short period time including knowledge creation, technology development, product production and process improvement derived from science and technology R&D activities. As the output directly related with R&D, the primary outcome includes patent, paper and prototype as well as research equipment.

The secondary outcome refers to social and economic effects derived from the primary outcome, including cost reduction, increase in sales and other social and economic contributions. As the succeeding outcome from the primary result, the secondary outcome is usually stated as a performance indicator associated with business activities such as market share, profit percentage, import substitution effects.

2.2 Management of the outcome of national R&D programs

The outcome derived from national R&D programs is a public funded important asset so that it should be properly managed to enhance its industrial economic utility. In fact, the ultimate goal of national R&D programs is to promote socioeconomic development through the use of the results from the R&D activities. Therefore, the outcomes of national R&D programs should be disclosed to all people and widely used through national management and distribution except for extraordinary cases. [4]. In particular, because the national R&D report, one of intangible research outcomes among national R&D programs, contains important matters on R&D such as R&D method and results, it is the final outcome of the national R&D program regardless of its type. Since it is the evaluation standard of the R&D project and major research outcome as the final outcome which takes place at the end of the R&D program, the systematic management of related information is especially important. [5]

Meanwhile, to prevent redundant research or investment and select a new research project through the mutual exchange of R&D information by informing various information derived from the national R&D programs to the general public and providing related information to researchers from the academia-industry-research cooperation institutes, there should be a system which can describe national R&D and manage general and detailed information. For the collection and supply of national R&D information, however, it is hard to develop a particular means but the national R&D report which includes the details and results of the R&D activities. Hence, it is critical to develop a management system which can and distribute manage national R&D reports in an integrated manner.

In this sense, the establishment of the information distribution system which makes it possible to collect, manage and use national R&D reports it recognized as one of the most important R&D policies in any country. In fact, each country has systematically managed related duties through a national science & technology information agency. After all, the management of national R&D reports is a critical part in the management of the research outcomes of national R&D programs.

3. Collection and management of national R&D report information

3.1 Progress and projects of national R&D programs

Korea's national R&D programs have continuously risen since the 1990s. For the last five years, they have grown by more than 10% annually. In 2012, for example, a total of KRW 15.9064 trillion was invested. The number of the R&D programs was as large as 537, and 43,192 projects were promoted, which means that the unit number of R&D projects is about 40,000.

<Table 1> Current Status of National R&D Programs, their Investments and Specific Projects by Year

Year / Category	2008	2009	2010	2011	2012
Investment (x KRW 100 million)	109,936	124,145	136,827	148,528	159,064
No. of Programs	486	474	483	493	537
No. of Specific Projects	37,449	39,565	39,254	41,619	43,192

※ Source: National Education Science & Technology Council,
National R&D Program Survey & Analysis Report, Annual Reference

3.2 National R&D report production trend

The national R&D reports (ex: final report, summary, etc.) derived from the national R&D programs have gradually increased. Since a strict survey on the number of final report started in 2012, it has declined. In principle, the final report of national R&D programs occurs by the unit of specific R&D project. Considering multi-year R&D programs and non-disclosure R&D projects, about 25% of the national R&D projects have been found.

<Table 2> National R&D Report Production Trend

Year / Category	2010	2011	2012	Total
No. of Report Produced	14,565	15,044	9,646	39,255

※ Source: The internal data (internal analysis) of the KISTI.

3.3 Current non-disclosure of national R&D reports by bureau

Among the R&D reports derived from the national R&D programs, constant ones can be kept undisclosed. In fact, a great portion of national R&D reports aren't open to the public. For the last three years, the reports' non-disclosure rate has been as high as 36.9%.

<Table 3> Classification of Non-disclosure Projects of National R&D Programs

Category	2010	2011	2012	Total (Average)
Total Projects	8,540	8,461	4,805	21,806
Disclosed Projects	4,399	5,635	3,736	13,770
Non-disclosure Projects	4,141	2,826	1,069	8,036
Non-disclosure Ratio (%)	48.5%	33.4%	22.2%	36.9%

※ Source: Ministry of Education, Science and Technology / KISTI, National Science and Technology Information System (NTIS)

According to analysis on current non-disclosure of national R&D reports, a big difference has been found among government bureaus. In terms of the percentage of non-disclosure when the Defense Acquisition Program Administration in which the particularity of national R&D programs was accepted was excluded, Small and Medium Business Administration was the highest with 100%, followed by Ministry of Knowledge Economy (63.3%), Ministry of Land, Transport and Maritime Affairs (32.1%), Ministry of Agriculture and Forestry (30.6%) and Ministry of Health and Welfare (22.3%).

<Table 4> Classification of Non-disclosure Projects of National R&D Programs by Bureau

	No. of Projects	Disclosed	Undisclosed	Percentage %
Ministry of Education, Science and Technology	8,622	8,302	320	3.7%
Ministry of Knowledge Economy	5,014	1,838	3,176	63.3%
Ministry of Agriculture and Forestry	471	327	144	30.6
Ministry of Health and welfare	780	606	174	22.3%
Ministry of Environment	388	371	17	4.4%
Ministry of Land, Transport and Maritime Affairs	308	209	99	32.1%
Small and Medium Business Administration	3,884	-	3,884	100%
Rural Development Administration	1,083	1,073	10	0.9%
Korea Food and Drug Administration	861	835	26	3.0
Defense Acquisition Program Administration	4	1	3	75.0
Total	21,415	13,562	7,853	36.7%

※ Source: Ministry of Education, Science and Technology / KISTI, National Science and Technology Information System (NTIS)

3.4 Collection and management of national R&D reports by the KISTI

For the government-led collection and management of national R&D reports, they should be submitted to or registered in the KISTI. However, the results have been very poor. In other words, the generation distribution of R&D information according to the related laws and the number of national R&D reports submitted to or registered in the KISTI have been very little. Since 2010, the registration standards for national R&D program reports changed to final reports. As a result, the volume of database on national R&D reports collected by the KISTI was 8,540 cases in 2010 and 8,461 cases in 2011. Based on the final reports, the percentage of database to the conventional national R&D reports was about 69%. With revision of the standards for the registration rate of national R&D program reports, the ratio of database was somewhat raised. However, the increase is very minor considering the functions and roles of the KISTI.

<Table 5> Development of National R&D Report Database by the KISTI

Year	2008	2009	2010	2011	2012
No. of Productions	17,233	18,160	14,565	15,044	9,646
No. of Database Development Cases	4,623	7,108	8,540	8,461	4,805

※ Source: Internal data of the KISTI

Among the national R&D reports collected by the KISTI, in particular, the percentage of non-disclosure national R&D reports is very high. According to the national R&D reports registered in the KISTI's NTIS, a total of 16,085 non-disclosure reports are registered as of March 2012 for the past five (5) years (2008 thru 2012), accounting for 48.0% of total national R&D reports. This kind of result means that when the reports are submitted or registered in the KISTI, those which are good to be disclosed are frequently kept undisclosed.

<Table 6> Collection of Non-disclosure National R&D Reports by KISTI

Category	2008	2009	2010	2011	2012	Total
No. of Total Reports Registered	4,623	7,108	8,540	8,461	4,805	33,537
No. Non-disclosure Reports	3,213	4,836	4,141	2,826	1,069	16,085
Percentage of Non-disclosure reports (%)	69.5%	68.0%	48.5%	33.4%	22.2%	48.0%

※ Source: Ministry of Education, Science and Technology /
KISTI, National Science and Technology Information System (NTIS)

4. Conclusion

As mentioned above, the information distribution systems for national R&D reports have been managed quite well. In other words, in addition to the information-linked distribution through the government bureau's information distribution system, the general registration system for national R&D reports has been operated from the perspective of the management of national R&D outcomes. From the formal aspect, therefore, the information management system of national R&D reports is well-designed and systematic.

However, current information management and distribution of national R&D reports are still very poor. As described above, the collection of the KISTI's national R&D reports and development of database on related information are way behind. Therefore, it is urgent for the government and institutions to come up with a solution.

Especially, it is critical to develop a system to disclose non-disclosure national R&D reports and distribute information. Among the national R&D reports, those which should be improved by the government or others requested by concerned parties such as participating firm can be kept undisclosed. Even so, the percentage of non-disclosure national R&D reports is significantly high. As a result, the collection of related non-disclosure reports by the KISTI is very low.

For the integrated registration and management and efficient distribution of national R&D reports, after all, there should be a more specific and practical plan. Then, the problems with which the reports are not properly submitted to or registered in the KISTI should be derived, and an improvement plan for them should be developed. For the systematic and unified submission and registration of national R&D reports, the use of a standardized R&D report form should also be suggested.

For the comprehensive distribution of the submitted or registered reports, moreover, an improvement plan for the related system should be developed for the efficient and effective development and operation of the government-led information distribution system by reviewing the related information together with a copyright issue which could be raised during the development of database and promotion of online services.

References

[1] Heo Jeon-oe, *A Study of Laws and Systems Relating to National Science & Technology Knowledge Information Infrastructure*, KISTI, Nov. 2002.

[2] Yoon Kwon-soon, *A Study of Current Ownership Regulations on Intellectual Property Rights in National R&D Programs and Improvement Plan, A Plan for Efficient Use of Intellectual Property Rights*, 26th JURIS FORUM records of Law Research Institute of Chungbuk National University, December 6, 2007.

[3] Yoon Jong-min, Chung Yong-il, Lee Seong-ho and Lee Hyung-jin, *A Study of Systems for Promotion of Information Distribution in National R&D Reports*, Journal of the 7th Korean Science & Technology Information Infrastructure Workshop, 2002.

Yoon Jong-min and Heo Jeon, *Legal Regulations on Public Technology Management*, Journal of Korea Technology Innovation Society, Vol. 9, No. 3, September 9, 2006.

[4] Refer to Article 16 (Improvement of National R&D Programs) of 'Regulations on the Management of National R&D Programs' (Presidential Decree No. 21634) and 'Regulations on Common Security Management of National R&D Programs (Ministry of Education, Science and Technology Decree No. 33).

[5] Heo Tae-sang, Lee Hong-ro and 5 others, *A Study of Management of National R&D Reports and Improvement Plan in Distribution Law*, Journal of 2009 Fall Symposium of Korea Technology Innovation Society, Korea Technology Innovation Society, November 27, 2009.

[6] Kim Seong-won, *A Review on National R&D Security Management System*, Journal of Korean Association for Industry Security, Vol. 1, No. 1.

[7] Park Jin-woo, *A Review on Non-disclosure Information According to Freedom of Information Act*, Dong-A Law, Vol. 43, 2009.

[8] Park Yeong-gil, *A Study of Promotion of Use of Digital Publications*, Ministry of Culture and Tourism, 2006.

[9] Chung Phil-woon, *Judgment on Possession of Government-created Copyrights and its Evaluation based on Constitutional Theories*, Korean Public Law Association, Studies on Public Law, book 38, Vol. 2, No. 1, October 2009.

[10] Yoon Jong-min, *A Study of Current National R&D Information Security Protection System and Tasks*, Journal of 2010 Fall Symposium of Korea Technology Innovation Society, 2010.

An Attempt to Nuance the Understanding of Professional Reports in Archaeology

Lisa Börjesson, Department of Archives, Libraries, Museums;
Uppsala University, Sweden

Abstract
Professional (i.e. extra-academic) contract archaeology is an internationally widespread practice contributing significantly to the archaeology literature. However, professional knowledge production in archaeology, and most notably the professional report genre, is at times described as problematic. The problem descriptions are ambiguous and can be grouped under at least three different topics: concerns for content quality and practical accessibility, concerns for the comparably low degree of analytical and theoretical synthesizing in reports and concerns for lack of mutual knowledge transfer between academic archaeologists and professional archaeologists. Technical issues of access are to an increasing extent being solved. Format standardizations are also developing. Hence the report genre becomes more accessible, and the content more readable and informative. Yet articulations of attitudes toward the genre in archaeology text books and journal articles remain focused on the genre's problems. The aim of my ongoing dissertation research is to nuance the understanding of the professional report genre in archaeology. I do so by analyzing factors shaping reporting as it takes place in the intersection between academic norms, professional values and market logics. I argue an improved genre understanding is crucial to diminish cultural issues of access to the report literature, and also as a basis for development of reporting practices. In the dissertation research I analyze (1) perceptions about the report genre in archaeology literature, (2) information policy regulating reporting in archaeology, (3) how report writers and county board professionals interpret the reporting and report auditing work tasks and (4) the frames of reference report writers bring into reporting. The aim of this paper is to explicate the research design consisting of four sub-studies, to briefly report on findings from study no. 4, and to discuss preliminary, partial results from study no. 2.

Keywords: *Documentation; Report; Professional report; Grey literature; Archaeology; Cultural heritage; Professional archaeology; Contract archaeology; Public government; Public administration; Cultural heritage management; Information policy; Science and technology studies; Scholarly communication; Sociology of professions; Mixed methods; Idea analysis; Policy analysis; Qualitative document analysis; Bibliometrics; Focus groups interviews*

1. Introduction

The rise and development of professional (i.e. non-academic) archaeology during the twentieth century has led to a situation where professional archaeology constitute a substantial part of all archaeological activities (Aitchison 2010; Kristiansen 1998; cf. Ambrosiani 2012). Professional contract archaeology has increasingly contributed to archaeology literature, primarily by producing reports from surveys and excavations undertaken prior to land development (Aitchison 2010; Ersgård 2006). Digital means for report production, dissemination, and archiving have made reports increasingly visible and accessible to users.

The professional report genre's role as a contribution to archaeological knowledge is subject to scrutiny (e.g. by Bahn 2012; Hodder 1989; cf. Lucas 2001). Discussions about the role of reports can be followed for example in two special issues of research journals around 2010: in *Archaeologies* from the viewpoint of archaeology and in *The Grey Journal* from the viewpoint of information science (both edited by the archaeologist Seymour 2010a; 2009a). These special issues cover topics like professional review (e. g. Harlan 2009), archiving and access (e. g. Kansa et al. 2010; Aitchison 2009), but also expound the relationship between professional archaeology and academic archaeology and the transfer of knowledge between the two (e. g. Aitchison 2010; Roth 2010; Seymour 2009b). Taken together the descriptions allude to a compound of issues where the problem focus at times lies within the genre, and at other times lies with attitudes towards the genre.

My ongoing dissertation research takes these ambiguous problem descriptions as a departure point and seeks to nuance the description of the archaeology report genre. The nuancing takes on the perspective of contract archaeology as a professional practice, embedded in heritage policy, administrative requirements, market logics, organizational conditions and professional ideals. The aim is, through an exploration of contextual factors shaping reporting to provide more informed grounds for readings and evaluations of reports. This re-contexualisation will serve as a basis to diminish some of the cultural issues of access to the genre, and from which to manage development of the report genre from policy and organizational levels. The overarching dissertation research questions are:

I. Which ideas about the report genre are articulated in discussions of 'the grey literature problem' in archaeology?
II. How are professional archaeology reports shaped by contextual factors such as heritage policy, market logics, administrative requirements, organizational conditions and professional ideals?
III. Can insights from studies of reporting contexts be used to address some of the issues articulated in the discussions of 'the grey literature problem'?

In this paper I lay out my mixed methods dissertation research design and report on findings from two of the four sub-studies. The paper begins by framing archaeological reporting as a genre of scholarly communication. The first section also explains my operationalization of the concept 'configuring factors' of documentation. The four sub-studies making up the dissertation research are described subsequently. The paper concludes by a discussion of the methodological approach and related ethical considerations on researching a professional discipline and its activities.

2. A genre and cultural issues of access

Throughout this paper I conceptualize archaeological reports as a genre. The use of the genre concept brings with it a grouping of individual documents. A library and information science (LIS) genre concept can also lead us to focus on the relations between documents, communication, activities and social organization (Andersen 2008, 339). Genres are assumed to not only be created in social settings, but to also create the settings by influencing structures for communication and actions (2008, 350). This view on the genre concept implies that reports shape archaeology, which motivates the dissertation research from an archaeology perspective.

If we begin with the understanding that the archaeology report genre is a boundary genre, dependent on both government heritage management and archaeology, we can assume the report genre is conditioned by multiple factors and therefore hard to describe once and for all (cf. Börjesson under review; Huvila 2011). Heritage management logics and archaeology develop both independently and interrelated and each make claims as to how contract archaeology documentation should be done. Furthermore the market organization of professional archaeology has brought about a multitude of organizational forms among the actors at the market, affecting the organizational contexts in which documentation is carried out (cf. Aitchison 2010).

Following the assumption that the report genre as a boundary genre is conditioned by multiple factors, it is likely that the multiple problems related to archaeological reporting not will be solved by any one measure. The genre's problems have been thoroughly problematized and argumentation pro and con the genre has been spelled out (Seymour 2010a; 2009a). The problems described are such as inadequate archiving, instable archives, survey data being disconnected from reports, perceptions about lack of peer-review, lack of interest in and respect of report content from academic archaeologists, and language barriers. I (tentatively) argue several of these problems arise when reports are read and evaluated in relation to academic ideals as a standard. One way forward could be, based on the assumptions that reporting is configured by multiple factors on the boundary between academic, governmental-professional and market logics, to seek further understanding of the non-academic factors shaping the reporting genre in professional archaeology. This approach would offer a model for understanding why reports become what they are and, by doing so, provide report readers with a

firmer ground from which to read and interpret report contents. Furthermore a model based on factors shaping professional archaeology reporting will provide a basis for evaluations and management of reporting practices.

The genre concept emphasizes common traits across documents. In order to nuance the understanding of the report genre I hence complement the genre perspective with the concept 'configuring factors' (of documentation) borrowed from Bernd Frohmann. Frohmann proposes a four-fold perspective to cover the "the configuring factors" of documents: the materiality of documents, their histories, the institutions in which they are embedded, and the social discipline shaping practices with them (Frohmann 2004, 405). This dissertation covers the last three mainly social factors, while leaving out the materiality. To explore configuring factors serves as a way to study factors shaping the genre. Variations within the genre are particularly interesting for a nuanced understanding of the genre. In my use of the term 'configuring factors' I therefore make the assumption that factors shaping a genre have unequal impacts different parts of the genre, resulting in variations of interest for overall understanding of the genre.

In a larger perspective this study can be placed in the field Science and Technology Studies (STS), particularly in the sub-field focusing on scientific (or 'scholarly') communication (Sismondo 2010). Professional archaeology reports are not simply scientific − they balance on the boundary between academia and government. Nevertheless information policy regulating the report genre expresses aspirations for the genre to be scholarly and to contribute to scientific knowledge (Riksantikvarieämbetet 2012). The boundary position is sometimes framed by the term 'grey literature' (Darvill 2008). Grey literature is a cross-disciplinary term used to describe literature produced on all levels of government, academics, business and industry where publishing is not the primary activity of the producing body (Farace and Schöpfel 2009).

3. Do we have a problem? Articulations of 'the grey literature problem' in archaeology literature (Study no. 1)

Articulations of the 'grey literature problem' in archaeology are ambiguous. They include descriptions of technical access challenges (e.g. reports are stored in regional depositories, Silva 2010), content usability issues (e.g. content is not validated by data, Roth 2010), and readers' confused expectations on the genre (e.g. false perceptions about lack of peer-review, Harlan 2010). These articulations point out deficiencies both within the genre and among its readers. A number of the voices also suggest ways to improve reporting and report use. Some of the articulations specify which national or other limitation of context they cover, yet other speaks about the problems in general terms.

This article analyzes general (mainly Anglo-American) articulations of the problems with the report genre in archaeology and how these problem articulations are reflected in Swedish archaeology literature. The aim is to explore nuances in the problem description and related 'solutions'. The guiding questions are: How are the challenges with grey literature described in the articles in the two special issues on professional archaeology documentation *Archaeologies* 2010 and *The Grey Journal* 2009? Are these articulations reflected in contemporary texts in Swedish? If so, what do the Swedish descriptions pick up and what do they leave out? The underlying assumption is that the problem descriptions not only describe a problem, but also foreshadow terms and forms for describing a phenomenon that is not necessarily in itself a problem. Hence the study does not seek to reinforce the problem, but to illuminate a cluster of ideas about the report genre in a reflective manner.

STS has a tradition of addressing problems or controversies. Controversies can be approached as a way to analyze a discipline at a point in history where conflicting perspectives intersect, are articulated and thus form rhetoric about phenomena, facts or artifacts. Another term used as a metaphor for this process is 'black boxing'. The STS study of controversies focuses at stages when a black box is open, when a phenomenon is about to be defined, or a previous dominating paradigm is being contested. As the controversy is settled the black box closes and the 'winning' argument become neutralized as 'knowledge'. The STS norm is to apply a symmetrical approach where all sides in conflict are analyzed as reasonable from their own perspective (Sismondo

2010). Important to note is that the ambiguous descriptions of the grey literature problem in archaeology does not necessarily qualify as a controversy in the sense that we see distinct opposite sides openly arguing for their own standpoint and against each other's. The grey literature problem in archaeology has a less outspoken character, but yet receives attention from a number of researchers and professional archaeologists from different angles (Seymour 2010a; Seymour 2009a).

The material will consist of articulations made by article authors in the two special issues of *Archaeologies* 2010 and *The Grey Journal* 2009, both on the topic grey literature in archaeology. The articulations made by the article authors will be compared to ideas in a selection of Swedish texts concerning the topic professional knowledge production and related challenges. The analysis method will be a qualitative document analysis (QDA) specifically focused on dimensions in the problem descriptions (Bergström and Boréus 2012; Plano Clark et al. 2010). The focus on dimensions in the problem descriptions is a way to encompass both the 'problems' and the 'solutions' written out or alluded to by the authors in the material.

4. Negotiating documentation in professional practices - The case of archaeology reporting (Study no. 2)

Information policy in heritage legislation and guidelines is the primary means by which the government can handle the report genre challenges. Secondary means for the government to use can be supporting academic heritage education and stimulating professional education among archaeologists and county administrative board professionals. Information policy comprises decisions, guidelines, regulations and laws directly involving information creation, processing, flows and use (Braman 2011). Information policy regarding reporting articulates ideas and norms on what reports should be like, hence how reports shall respond to the needs of all of the stakeholders whose voices have been heard in the policymaking process.

Heritage policy is comprised of both hard law on for example, heritage interests and property rights and on international antiquities trade, and soft law on how surveys should be documented and how documentation subsequently should be handled. The hard law portions of heritage legislation has assigned enforcement institutions such as the property and environment court in the case of property rights and the border police in the case of antiquities trade. Enforcement of soft law is based upon non-legally binding regulations and guidelines by government authorities and non-governmental organizations in a field. There is also a moral responsibility resting on archaeology and heritage management professionals, and on professional associations to act according to regulations.

Preliminary results from an analysis of information policy in Swedish cultural heritage policies show a low level of regulation of archaeology reporting in the heritage legislation. The county administrative boards are assigned the juridical responsibility to ensure adequate documentation and reporting, and also to set standards for what is being seen as 'adequate'. County administrative boards request reporting in public tenders, but also require archaeology contractors to establish documentation plans. The practical governance model hence distributes the responsibility between the county boards as clients and auditors, and the archaeological contractors as executors (Börjesson, Petersson, and Huvila work in progress). A further analysis will focus on the level of county administrative boards and archaeological contractors. The aim of the analysis is to explore how policies on reporting and professional responsibility play out in negotiations between public procurement tenders and archaeology contractors' documentation plans. QDA will be applied to structure the analysis and to pursue a discussion of how the ideas from the national level information policy play out in the interactions between county administrative boards and archaeological organizations in the case of documentation plans (Bergström and Boréus 2012; Nyqvist 2011; Cris Shore and Wright 2011; Plano Clark et al. 2010).

5. Professional interpretations of the reporting and report auditing work tasks – The case of professional archaeology (Study no. 3)

The third study explores archaeology professionals' interpretations of the reporting work task and government authority professionals' interpretations of the report auditing work task. While

previous studies tracks the articulations of the grey literature problem in archaeology and information policy concerning reporting this study analyzes how reporting archaeologists and those overseeing the reporting practice formulate their view on their respective work task. The aim is to find out how professionals interpret reporting and report auditing as work tasks and responsibilities. The analysis will provide a basis for discussing professionals' points of view on reporting.

The material for this study will consist of results from previous studies of reporting (primarily government reports) and qualitative focus group interviews with county administrative board professionals and with professional archaeologists (at separate occasions). The two types of professionals will be given a similar set of points of discussion, but slanted toward their practices at 'opposite ends' of the reporting process. Results from the previous sub-studies may be incorporated into the material for the interviews.

The interview records will be analyzed in the light of theory of professions (Svensson and Evetts 2010). The key assumption upon which the choice of theory is based is that reporting is made in a primarily professional environment, outside academia and embedded in an organizational setting with more shaping forces like topical training, professional ideals and obligations toward organization and colleagues etcetera, which interacts with the policy regulating reporting. Possible theoretical terms to operationalize in the interpretation are 'civic epistemologies', 'professional (or 'expert') knowledge' and 'information interests'
(Mosse 2011; Jasanoff 2005; Sundin and Hedman 2005).

6. Grey literature – grey sources? Report writers' frames of reference (Study no. 4)

In this study I explore report writers' frames of reference in order to understand more about variations among report writers' perspectives (Börjesson under review).[1] Frames of references are studied by an analysis of source use patterns in bibliographical reports. The source use patterns are analyzed based on a detailed coding of cited sources in 97 Swedish field evaluation reports from 2013. The coding structure was developed iteratively to cover aspects represented in bibliographical lists. The variables catch source age, source type, source format, source language, organizational and spatial relation between source originator and report author.

Information source use in archaeology in general has been studied previously (Huvila 2014a; Huvila 2006). Use of specific types of sources has also received attention, specifically image use (Beaudoin 2014; Moser 2012). Isto Huvila's study target variations in source use among archaeologists related to work roles. My study focuses specifically on source use in professional reports and explores variations in source use across the variables through a correspondence analysis. The bibliometric approach provides means to analyze distributions within the dataset, to test connections between parts of data and to reveal latent patterns (Denscombe 2009, 327; Tague-Sutcliffe 1992, 1).

The study begins with the assumption that as archaeological reports primarily are written in extra-academic, professional settings, it is relevant to seek to understand report writers' frames of references in relation to professional standards colored by organizational affiliations (Kristiansen 1998). Organizational types are therefore a key aspect in the analysis and interpretation, aided by sociology of professions influenced LIS research discussing information cultures and information interests shaped by occupational identities (Widén-Wulff 2010; Sundin and Hedman 2005).

The findings partly confirm previous studies of information source use, but also show some notable discrepancies. Reports are a key source for report writers in addition to professional literature and maps. Notwithstanding maps, non-codex like information sources are rare in the bibliographical lists. In previous studies both archaeological materials, aggregates of original materials and images have been described as information sources vital to archaeologists (Beaudoin, 2014; Huvila, in press a; cf. Moser, 2012). The complete lack of references to archaeological materials and the very rare references to images suggest that report writers either

[1] The dataset will be made available along with the article upon acceptance for publication.

do not use these sources or that there are circumstances hindering archaeologists from representing these materials as sources. The national character of source choices shown in previous studies is largely confirmed. The source age distribution analysis shows a strong correlation between source type and source age. The importance of novelty is in this case related to source type and also possibly to additional contextual issues as accessibility.

References to databases and webpages (not including map databases) are about as common as references to academic literature, like the results in Huvila's study (2014a). Two notable characteristics of the use of digital sources are the relatively few digital sources, and the very rare occurrences of references to sources other than codex-like sources. Digitization and born-digital information in other forms such as digital dataset, objects in digital collections, 3D models, virtual reality visualizations seem to be rarely or never represented.

The correspondence analysis used as a way to analyze source use patterns reveal one administrative, one professional/academic pattern, and one map pattern. These three clusters show that the main division between source use patterns in the report genre is the division between the administrative source use pattern and the professional/academic source use pattern. This finding implies that report authors primarily relate to two spheres of previous knowledge production: an administrative sphere containing planning documents and reports more closely related to local cultural heritage management and a professional/academic sphere comprising more of the type of sources traditionally recognized as academic, but also a range of non-academic professional publications. These two patterns suggest a variation within the report genre, where the dividing line goes between administrative and professional/academic source use patterns.

This study contributes by an examination of archaeological information use in a specific context, which has been requested in previous research (Huvila, in press a; Moser, 2012; Kansa and Kansa, 2011). The results can be employed to nuance both the understanding of professional archaeologists' information use, and to refine the understanding of the report genre from an information source use point of view. Awareness of variations between frames of reference among report writers could alter the perception of reports as "grey literature", in the sense being not academic (cf. Seymour 2010b).

7. Researching factors shaping scholarly communication in the intersection between academia, government and market – A brief discussion of methodology and research ethics

Above I make the argument that professional reporting in archaeology is influenced by a range of factors and that these factors, at least to some extent, are overlooked in evaluations of the report genre. Within the scope of the dissertation research I explore four of these factors; ideas in articulations of the genre's problems, policy regulating the genre, report writers' and county board administrators' interpretations of the reporting and report auditing work tasks, and report writers' frames of reference. Borrowing an anthropological metaphor, these four factors can be seen as 'sites' at the professional reporting field (cf. Wright 2011). From that point of view the dissertation research become a ethnography of a professional documentation practice with the aim to nuance the understanding of the genre produced within the practice.

As for ethnographic research in general the framing of the research problem and the choice of sites from which to study the problem is pivotal. In the case of my research the frame of the research problem is the report genre challenges as expressed by archaeologists. The description of these challenge articulations and how general articulations travel to and reflect in the Swedish situation will be further detailed by study 1. The choice of sites, apart from the first case study illuminating the challenge articulations, follows a stylized timeline in the professional archaeological process: from heritage policy through professionals' interpretation of the work task and finally implementation. Moreover the choice of sites cover a macro-micro dimension; heritage policy regulate all archaeologists, interpretations of the reporting and report auditing work tasks are influenced by the policy level but mainly take form in local social interactions and

finally the report writers' frames of references reflect the perceptions of previous knowledge individuals bring with them into report writing.

The ethical considerations impacting this research circulate around the notion of professional practice. Professionals at work have reasons for the way they do their work which should be acknowledged in the analysis. However these reasons may not always be flattering, and may collide with strivings for a professional identity (cf. Mosse 2011). Ethnographic researchers must be prepared to balance the goal of ethnographic research with a respect for those studied. Furthermore the stylized timeline on which this dissertation research is based cannot be seen as a mirror of the practice, but more of as a model upon which the public sector is structured (from policy to professional interpretation and through to implementation), and thus from which public sector processes can be analyzed. This type of ethnographic multi-sited research will never explain individual events, but seeks to understand a phenomenon on a higher degree of abstraction. Ethnographic research traditionally attempts to let the studied community have an opportunity to review the research, through some type of public dissemination (Mosse 2011). It is less common that ethnographic research seeks to arrive at suggestions for changing the particular practice studied. This study should plan for a communication with the practice studied, and should in my opinion also consider if there are practical suggestions to make based on the thesis research regarding management of reporting practices.

References

Aitchison, Kenneth. 2009. "Standards and Guidance in Archaeological Archiving: The Work of the Archaeological Archives Forum and the Institute for Archaeologists" *The Grey Journal* 5 (2): 67–71.

———2010. "Grey Literature, Academic Engagement, and Preservation by Understanding" *Archaeologies* 6 (2): 289–300. doi:10.1007/s11759-010-9145-5.

Ambrosiani, Björn. 2012. "Swedish Contract Archaeology in the 1950s and 1960s" In *Histories of Archaeological Practices*, edited by Ola W. Jensen, 305–24. Stockholm: Statens Historiska Museum.

Andersen, Jack. 2008. "The Concept of Genre in Information Studies" In *Annual Review of Information Science and Technology*, 339–67. Medford, NJ: Information Today.

Bahn, Paul G. 2012. *Archaeology: A Very Short Introduction*. Oxford: Oxford University Press.

Beaudoin, Joan E. 2014. "A Framework of Image Use among Archaeologists, Architects, Art Historians and Artists" *Journal of Documentation* 70 (1): 119–47.

Bergström, Göran and Kristina Boréus. 2012. "Idé- och ideologianalys" In *Textens mening och makt*, edited by Göran Bergström and Kristina Boréus, 149–80. Lund: Studentlitteratur.

Börjesson, Lisa. under review. "Grey Literature - Grey Sources? Nuancing the Perception of Professional Documentation: The Case of Archaeology"

Börjesson, Lisa, Bodil Petersson and Isto Huvila. work in progress. "Information policy for sustainable digitization in archaeology? – A critical analysis"

Braman, Sandra. 2011. "An Introduction to Information Policy." *Information Policy*, no. 1: 1–5.

Darvill, Timothy. 2008. *The Concise Oxford Dictionary of Archaeology*. Oxford: Oxford University Press.

Denscombe, Martyn. 2009. *Forskningshandboken*. Lund: Studentlitteratur.

Ersgård, Lars. 2006. "Det uppdragsarkeologiska alternativet" In *Det förflutna är att räkna med*, edited by Lars Ersgård, 10–25. Stockholm: Riksantikvarieämbetet.

Farace, Dominic J. and Joachim Schöpfel. 2009. "Grey Literature" In *Encyclopedia of Library and Information Sciences, Third Edition*, 2029–39. Oxford: Taylor & Francis.

Frohmann, Bernd. 2004. "Documentation Redux: Prolegomenon to (Another) Philosophy of Information" *Library Trends* 52 (3): 387–407.

Harlan, Mark E. 2009. "Archaeology's Grey Literature and Professional Review" *The Grey Journal* 5 (2): 83–88.

———.2010. "Black and White Literature, Grey Scholars." *Archaeologies* 6 (2): 270–88. doi:10.1007/s11759-010-9143-7.

Hodder, Ian. 1989. "Writing Archaeology: Site Reports in Context" *Antiquity* 63: 268–74.

Huvila, Isto. 2006. *The Ecology of Information Work: A Case Study of Bridging Archaeological Work and Virtual Reality Based Knowledge Organisation*. Åbo: Åbo akademis förlag.

———.2011. "The Politics of Boundary Objects: Hegemonic Interventions and the Making of a Document" *Journal of the American Society for Information Science and Technology* 62 (12): 2528–39.

———.2014a. "Archaeologists and Their Information Sources" In *Perspectives to Archaeological Information in the Digital Society*, edited by Isto Huvila, Meddelanden Från Institutionen För ABM. Uppsala: Department of ALM, Uppsala University.

———.2014b. "Introduction" In *Perspectives to Archaeological Information in the Digital Society*, edited by Isto Huvila, Meddelanden Från Institutionen För ABM. Uppsala: Department of ALM, Uppsala University.

Jasanoff, Sheila. 2005. *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton, N.J.: Princeton University Press.

Kansa, Eric C. and Sarah Whitcher Kansa. 2011. "Toward a Do-It-Yourself Cyberinfrastructure: Open Data, Incentives, and Reducing Costs and Complexities of Data Sharing" In *Archaeology 2.0*, edited by Eric C. Kansa, Sara Whitcher Kansa and Ethan Watarall, 57-92. http://escholarship.org/uc/item/1r6137tb#page-73.

Kansa, Eric C. et. al. 2010. "Googling the Grey: Open Data, Web Services, and Semantics" *Archaeologies* 6 (2): 301–26. doi:10.1007/s11759-010-9146-4.

Kristiansen, Kristian. 1998. "Gamla gränser och nya frontlinjer" In *Arkeologiska horisonter*, edited by Håkan Karlsson, 176–99. Stockholm: Symposion.

Lucas, Gavin. 2001. *Critical Approaches to Fieldwork: Contemporary and Historical Archaeological Practice*. London: Routledge.

Moser, Stephanie. 2012. "Archaeological Visualization - Early Artifact Illustration and the Birth of the Archaeological Image" In *Archaeological Theory Today*, edited by Ian Hodder, 292–322. Cambridge: Polity.

Mosse, David. 2011. "Politics and Ethics: Ethnographies of Expert Knowledge and Professional Identities" In *Policy Worlds: Anthropology and the Analysis of Contemporary Power*, edited by Cris Shore, Susan Wright and Davide Però, 50–67. New York: Berghahn Books.

Nyqvist, Anette. 2011. "Sweden's National Pension System as a Political Technology" In *Policy Worlds: Anthropology and the Analysis of Contemporary Power*, edited by Chris Shore, Susan Wright and Davide Però, 205–22. New York: Berghahn Books.

Plano Clark, Vicki L. et. al. 2010. "Mixing Quantitative and Qualitative Approaches - An Introduction to Emergent Mixed Methods Research." In *Handbook of Emergent Methods*, edited by Sharlene Nagy Hesse-Bieber and Patricia Levy, 363–87. New York/London: The Guilford Press.

Riksantikvarieämbetet. 2012. *Uppdragsarkeologi (2 Kap, 10-13 §§). Vägledning för tillämpning av kulturminneslagen*. Stockholm: Riksantikvarieämbetet. http://www.raa.se/wp-content/uploads/2012/06/V%C3%A4gledning-Uppdragsarkeologi-avsnitt-1-9.pdf.

Roth, Barbara J. 2010. "An Academic Perspective on Grey Literature" *Archaeologies* 6 (2): 337–45. doi:10.1007/s11759-010-9141-9.

Seymour, Deni J. ed. 2009a. "The Grey Journal - Archaeology and Grey Literature" 5 (2).

———.2009b. "Special Ethical and Scholarship Considerations for the Archaeological Grey Literature." *The Grey Journal* 5 (2): 89–96.

———ed. 2010a. "Archaeologies - Black-and-White Issues About the Archaeological Grey Literature," no. 2.

———.2010b. "In the Trenches Around the Ivory Tower: Introduction to Black-and-White Issues About the Grey Literature." *Archaeologies: Journal of the World Archaeological Congress* 6 (2): 226–32.

Shore, Cris, and Susan Wright. 2011. "Conceptualising Policy: Technologies of Governance and the Politics of Visibility" In *Policy Worlds: Anthropology and the Analysis of Contemporary Power*, edited by Cris Shore, Susan Wright, and Davide Però, 1–26. New York: Berghahn Books.

Silva, Rosiclér Theodoro da. 2010. "Some Aspects of 'Grey Literature' in Brazil" *Archaeologies* 6 (2): 327–36. doi:10.1007/s11759-010-9142-8.

Sismondo, Sergio. 2010. *An Introduction to Science and Technology Studies*. Chichester, West Sussex, U.K: Wiley-Blackwell.

Sundin, Olof, and Jenny Hedman. 2005. "Theory of Professions and Occupational Identities." *Information Today*, 293–97.

Svensson, Lennart G. and Julia Evetts. 2010. *Sociology of Professions: Continental and Anglo-Saxon Traditions*. Göteborg: Daidalos.

Tague-Sutcliffe, Jean. 1992. "An Introduction to Infometrics" *Information Processing & Management* 28 (1): 1–3.

Widén-Wulff, Gunilla. 2010. "Organizational Culture." In *Encyclopedia of Library and Information Sciences, Third Edition*, 4073–78. Oxford: Taylor & Francis.

Wright, Susan. 2011. "Studying Policy: Methods, Paradigms, Perspectives" In *Policy Worlds: Anthropology and the Analysis of Contemporary Power*, 27–31. New York: Berghahn Books.

Open Access Korea, Phase 1: Five Years On

Hyekyong Hwang, Seonheui Choi, Hyesun Kim;
Korea Institute of Science and Technology Information, KISTI, Korea

Introduction

As the world began to take interest in the open access, the Korean government has been taking an interest in the open access as well, and policies in relation to information sharing are being development. The Ministry of Culture, Sports and Tourism established a comprehensive plan for library development, and open access movement started in 2009 for the sharing and spreading the research output published in Korea. A Korean style repository was developed and supplied as a technical foundation to realize the open access, and an electronic publishing system started being established for the open access journal publishing. Moreover, in order to promote policy/system strategies for the open access of the research output produced by Korean researchers, public access policies have been studied and expert groups have been formed and forums have been held. Various types of workshops and conferences were held as well in order to raise awareness of open access in Korea. This paper introduce such projects and activities of OAK for the past five years and various effects on policy, system and service as well as future tasks of OAK.

Project Outcomes and Activities

OAK Institutional Repositories(OAK IR) and OAK Portal

As a part of the OAK project, the Korea Institute of Science and Technology Information and Argonet Company developed a Korean open access repository system. OAK IR was developed as a part of a green road strategy for the spreading of OA. OAK IR is a repository system customized to meet the needs of Korean circumstances based on an open source, Dspace, developed by the Massachusetts Institute of Technology and HP. This system was formed with the institutional customizing module, the Korean common module, and the global base module.

- Institutional Customizing Module: Providing website UI of OAK IR, linkage with the internal management system (e.g. digital library system, research management system)
- Korean Common Module: Providing Hangul morpheme analyzer, automatic system for key word search , statistics system
- Global Base Module: Registration into global site (OpenDOAR, ROAR), Open API

Open Access is not commonly known in Korea. Thus, strategies were established to select organizations that are willing to actively share the knowledge and information of the organisation and to open access the collected contents, and to install OAK IR. Every year, applications were received and a group of expert evaluators evaluated organizations to which OAK IR is to be provided. The selected organizations were visited by system engineers, and the engineers analyzed, researched the demand, and installed the OAK IR. The selections were made based on evaluation of three categories—organization, systems, and open access contents.

- Organization: Willingness of the representative to operate the repository, a department and personnel to be solely responsible for the IR operation
- System (H/W, S/W): Whether the organization satisfies the minimum basic conditions for operation of the repository
- Open access contents: Diversity, richness and level of academic professionalism of the contents that are able to be openly accessed

25 organizations were selected after evaluation of the above factors by the expert evaluators. The types of these organizations range from public organizations, research centers, university libraries and public libraries. About 300,000 contents are currently contained in OAK IR, including academic articles, reports, patent documents, etc. The OAK IR system developed in Korea is now publicly accessible as open source, and can be downloaded and installed for free on OAK Portal.

OAK Portal is a comprehensive search platform for open access contents in Korea(refer to <figure 1>). The contents provided by OAK Portal include metadata of contents established and operated by individual organizations as well as contents contained in OAK IR. OAK Portal regularly collects metadata from individual IR through Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), the international standard for metadata exchange, indexes and provides a comprehensive search service.

In particular, the contents metadata collected from each organization goes through a process of filtering by way of contents quality management before they are provided. The contents quality management process of OAK Portal is as follows.

The items of metadata collected in the OAK Portal are categorized into mandatory/ optional, and the categorized items in the OAK Portal are regularly examined. Through this process, error data are discovered, and managers of OAK Portal implemented a system of recollecting the error items in the OAK Repository of each organization.

Through such process, materials such as academic papers, RnD reports, patents, multimedia contained in IR of each organization are collected, and collected information are provided through a browsing service by each organization, type of information and subject.

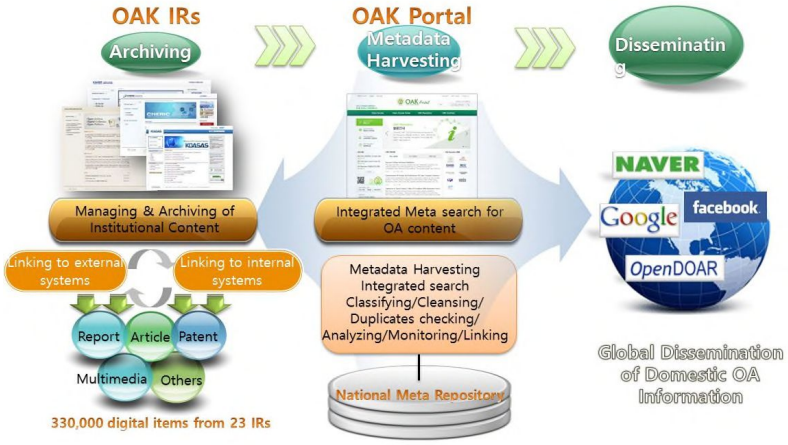


Figure 1. IR content archiving and dissemination process

Open Access Journal and OAK Central

World renowned publishers such as Springer tend to convert PDF files into XML files in order for the academic papers to become more accessible and readable. A global representative medical archive, PMC (Pubmed Central) provides a PMC Journal Publishing DTD for production of high-quality academic papers.

As Korean academic societies had been providing academic journals on the internet for free, they did not know the difference between open access and free access. Therefore, the project researchers had difficulties to explain the concept and necessity of open access to the executive members of academic societies

In this regard, in this project, we have thought about strategies to induce publishing of open access academic journals in Korea. While the awareness level of academic societies for open access journals was low, we have proposed free establishment of full texts into XML for the academic papers in order to persuade the editors. The editors had strong demand to make their academic society journal into SCI level journal. Having understood this need, this OAK project attempted to induce conversion of the selected academic papers according to JATS 1.0 DTD into full text XML for free.

As a result, 25 titles of academic journals and 2,979 articles in Korea have been established into full-text XML, and the academic journals published in the academic societies have been able to change into open access journal. KISTI produced a guideline to establish full-text XML and established articles in a standardized form, verified XML structuring, using the PMC XML validation checker, and produced open access/free access information in 25 titles of academic journals.

Moreover, in order to convert the electronic versions (PDF) of academic papers into XML format, automation function was developed in addition to XML Workbench, which enabled shortening of time required for establishment of each academic paper.

OAK Central is a platform to archive Korean open access journals and provide them globally(refer to <figure 2>). Through this website, Korean open access academic papers are able to be searched by table, figure, and ppt. In addition, this website also provides a function to view related articles, and data extract functions based on OAI-PMH. This is the result of OAK project research team in objectifying Korean academic papers by analyzing Journal Publishing DTD 3.0 developed by NCBI.

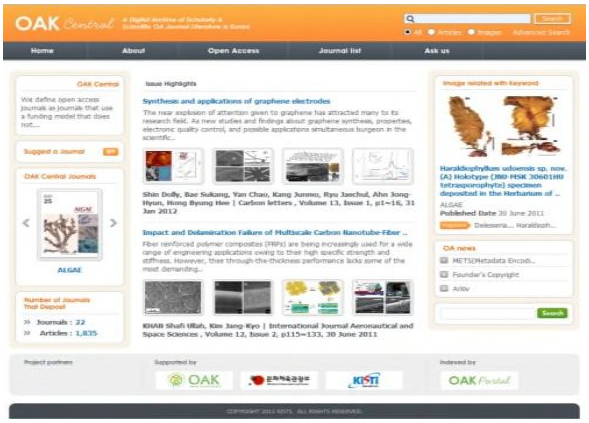


Figure 2. OAK Central(http://central.oak.go.kr)

Research on Public Access Policy

In 2010, through research of public access policies the publicly funded research output, domestic and international rules and regulations were analyzed and public access policy proposals were developed. And in 2011, laws and policies related to public research outputs were analyzed, and regulations and processes were developed. With the drafts of law, public hearings were held with organizations in related fields, but it was difficult to gather opinions from the participating stakeholders. Awareness of open access, and the conditions for acceptance from the academia, libraries, publishers, etc. were deemed inadequate to finalize the drafts of law. It would be necessary to gather opinions with stakeholders through discussions in order to enact drafts of law related to public access policies in the future.

To gather such opinions, OAK project team and the government need to continue to promote OA with strong will. Furthermore, mid and long-term plans, such as funding support, introduction of incentive systems, would be necessary. However, there are some issues to be faced for institutionalization of the open access of research outputs with public funds.

First, Korea is managing research outputs of government RnD project, but the scopes of research outputs are limited to RnD reports, and there are no regulations on public access of academic journal papers, which are the key points of academic communication. It would be necessary to prepare public access policies to include such academic journal papers.

Second, the awareness of open access is low for Korean researchers, and the publicly funded research outputs are perceived as private property of the researchers, it is difficult to arrive at an agreement of opinions on open access.

Third, Korean DB distributors strongly resist academic society journals being changed into open access journals, concerned with reduction of income from the sales of academic database.

Other Promotional Programs and Activities

Through this project, we have conducted various activities for vitalization of OA in Korea.

First, we operated forums for open access expert and open access academic journal editors.

Second, for the development of open access business model, we studied the cooperation scenario with SCOAP3 and KESLI, a Korean consortium for electronic journals.

Third, to raise awareness of Koreans, we have produced and disseminated online various types of education materials and promotion videos.

Fourth, we have produced a guideline to operate an open access repository through the special group for OAK repository managers and held many types of workshops to share management know-how.

Fifth, by holding OAK conferences in Korean during the open access week, we tried to contribute to raising awareness of Korean researchers. The OAK project team promoted scholarly communication of Korea to governments, foundations, organizations, libraries, researchers and users through such conferences.

The purpose of these conferences was to make the stakeholders related to scholarly communication a group in favour of open access.

Discussions

While proceeding with OAK project for the past 5 years, we have tried to raise awareness of domestic researchers and members of academic societies. In particular, resistance from DB producers who had established and sold academic paper DBs was very strong, and affected open access movement.

Scholarly communication in Korea is seldom commercial, which is different from overseas. Academic journals are published mainly by academic societies, and the costs for publishing academic societies consist of membership fees from academic society members and the government funds. However, not all of the academic societies are funded by the government, academic journals that have been evaluated as excellent have been supported by the government.

In this regard, the Korean Federation of Science and Technology Societies, which supports funds of scientific, technological academic journals, and National Research Foundation of Korea, which funds academic journals on liberal arts, social science and science, began to include "whether they are openly accessible" in the factors to be evaluated. This is the result of raising awareness of open access in Korea through OAK project, and the government took this awareness into effect by systemization.

Moreover, another effect of this project is that public fund management organizations such as Korea Centers for Disease Control & Prevention have started to manage public access policies on research outputs.

In the case of OAK Repository, the number of repositories is increasing, but there are still issues on system and management to be complemented. Especially, it would be necessary to codify policies and/or management guidelines on the point that the research outputs created by researchers who belong to organizations should be required to be deposited and openly accessed.

In the case of Korea, repositories are operated based on libraries that belong to organizations, cooperation with related departments (IT and research output departments) continues to be needed.

This project, upon completion of the research and development for five years, is proceeding to continue on with Phase 2 with the National Library of Korea, and will continually try to solve such remaining issues.

References

OAK(Open Access Korea) Portal http://www.oak.go.kr

OAK Central http://central.oak.go.kr

Open Access Korea, Phase 1, the final report. [internal document].

Think Tanks, Twitter, and Grey Literature

Danielle Aloia
The New York Academy of Medicine, NYAM, United States

Abstract
This poster aims to track the use of twitter among think tanks and the dissemination of grey literature.Making connections between the future of the online world and knowledge communication to drive change and share resources.

Findings:
In my daily work I collect and catalog grey literature in public health. The field of public health is quite large and I was looking for a tool that would help me to find materials in an easier way than going from website to website. I began following organizations on Twitter, like World Health Organization and the Robert Wood Johnson Foundation. Low and behold, I was collecting 100s of reports a day and realized this was a good way to disseminate information, especially grey literature. I came across the article Translating Research for Health Policy: Researchers' Perceptions and Use of Social Media in *Health Affairs* and it sparked an interest in doing more research to see what was being researched about this topic. To my surprise there was quite a bit being written. I decided to focus on Twitter because it is a tool for immediate access to what is going at any given moment of the day.
Many think tanks are using social media to connect with their followers by sending a quick tweet with links to longer reports or webpages. Twitter is a direct-to-direct communication channel that allows researchers and organizations to engage with each other and their audiences. Twitter is used to communicate scientific evidence, to keep up with what is going on in areas of interest, to share ideas, and to gain visibility. As the Brookings Communication Advisor, David Jackson states: "The opportunities to engage are opening up. If it's something in the news that you care about, you can quickly talk about it and you want to establish yourself as a resource so when the report comes out, people will think of you."[1]
Researchers themselves are being encouraged to use this resource to share ideas and build communities. Indeed, dissemination of grey literature should be just as important as the production of it.[2] Some reservations and concerns expressed in using Twitter are: it is not seen as relevant, may scar credibility, is a waste of time, not reaching targeted audiences.[3] In order to ensure the proper use of this tool researchers and think tanks want proven strategies for effectiveness and would like clear guidelines on how to use it effectively. In a webinar given by AcademyHealth, they suggested using Twitter to: summarize findings, general dissemination, visibility, and to test ideas and not be afraid to make mistakes.[4]
Twitter is, also, being scanned and analyzed to evaluate the effectiveness, reach, and impact of think tanks. Followers retweet tweets that are of interest to them. According to P. T. Metaxas et al. "retweeting indicates, not only interest in a message but also trust in the message and the originator, and agreement with the message content."[5] In the US, the Heritage Foundation and Cato Institute dominate social media and web traffic.[6] Thirty percent of all traffic to think tanks comes from Google and Facebook.[7] So, it is a definite advantage to be at least on one outlet of social media.
Don't be afraid to repeat your tweet: *Repeating a tweet can really save your news if you had poorly timed it the first time.*[8] Facebook may have more followers than Twitter, but Twitter users have wider networks, and according to a Pew Research poll some 46% use Twitter daily, with 29% checking in several times per day.[9] Remember Facebook is for people you know and Twitter is for people you want to know.
Today, it is important to keep up with all media outlets because in the future the ability to access information will be embedded in everything we do. These new communication channels will help disseminate grey literature and provide direct contact with the people making policy decisions. In this way research that may otherwise have been unseen will have a wider audience and a better chance at shaping policy-making in the future.

References

[1] Barbara Ray. How Think Tanks Are Using Social Media. *HiredPen Inc.* Available at: http://www.hiredpeninc.com/blog/how-think-tanks-are-using-social-media/. Accessed October 27, 2014.

[2] Danielle M. Cossarini, Bertrum H. MacDonald, Peter G. Wells. Communicating marine environmental information to decision makers: Enablers and barriers to use of publications (grey literature) of the Gulf of Maine Council on the Marine Environment. *Ocean & Coastal Management* 2014;96:163-172. doi:http://www.sciencedirect.com/science/article/pii/S096456911400163X.

[3] David Grande, Sarah E. Gollus, Maximilian Pany, et al. Translating Research for Health Policy: Researchers' Perceptions And Use Of Social Media. *Health Affairs* 2014;33(7):1278-1285. doi:10.1377/hlthaff.2014.0300.

[4] Holve E, Bhattacharya N, Savitz L, Trotter F, Wu A. Why Tweet? Building the Skills to Engage New Audiences and Promote Your Work. 2014. Available at: http://www.academyhealth.org/Training/ResourceDetail.cfm?ItemNumber=14693.

[5] Metaxas PT, Mustafaraj E, Wong K, Zeng L., O'Keefe M, Finn S. Do Retweets indicate Interest, Trust, Agreement? (Extended Abstract). 2014. Available at: http://arxiv.org/pdf/1411.3555v1.pdf.

[6] Julia Clark, David Roodman. *Measuring Think Tank Performance: An Index of Public Profile*. Washington, DC: Center for Global Development; 2013:33. Available at: http://international.cgdev.org/sites/default/files/think-tank-index_0_0.pdf.

[7] Alejandro Chafuen. Free Market Think Tanks: Website And Social Media Presence - Forbes. *Forbes* 2014. Available at: http://www.forbes.com/sites/alejandrochafuen/2014/02/12/free-market-think-tanks-website-and-social-media-presence/. Accessed October 27, 2014.

[8] Benjamin Rey. Should you repeat your tweets? Facts and figures from 1M tweets. *Wisemetrics* 2014. Available at: http://blog.wisemetrics.com/should-you-repeat-your-tweets-facts-and-figures-from-1m-tweets/?utm_content=buffer84aba&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer.

[9] Duggan M, Smith A. *Social Media Update 2013: 42% of Online Adults Use Multiple Social Networking Sites, but Facebook Remains the Platform of Choice*. Washington, DC: Pew Research Center; 2013:18. Available at: http://www.pewinternet.org/files/2013/12/PIP_Social-Networking-2013.pdf.

Marine Planning and Service Platform (MAPS) An Advanced Research Engine for Grey Literature in Marine Science

Sara Goggi, Monica Monachini, Francesca Frontini,
Roberto Bartolini, Gabriella Pardelli, ILC-CNR, Italy;
Maurizio De Mattei, Franco Bustaffa, DP2000, La Spezia, Italy and
Giuseppe Manzella, ETTsolutions, Genova, Italy

Abstract
*The MAPS (Marine Planning and Service Platform) project is a development of the **Marine** project (Ricerca Industriale e Sviluppo Sperimentale Regione Liguria 2007-2013) aiming at building a computer platform for supporting a Marine Information and Knowledge System, as part of the data management activities. One of the main objective of the project is to develop a repository that should gather, classify and structure marine scientific literature and data thus guaranteeing their accessibility to researchers and institutions by means of standard protocols.*
We will present the scenario of the Operative Oceanography together with the technologies used to develop an advanced search engine which aims at providing rapid and efficient access to a Digital Library of oceanographic data. The case-study is also highlighting how the retrieval of grey literature from this specific marine community could be reproduced for similar communities as well, thus revealing the great impact that the processing, re-use as well as application of grey data have on societal needs/problems and their answers.

1. Introduction

Operational Oceanography is the branch of marine research which deals with the development of integrated systems for monitoring, analyzing, modeling and forecasting oceans and seas. The integrated systems need access to real-time as well as delayed mode data. The operational oceanography activities includes also the ri-analysis of data and models to assess marine variability or long term trends. These activities require the access to quality controlled data and to information that is provided in reports and/or in relevant scientific literature. This finds application in many areas, ranging from environmental studies, security and safety to protection of off-shore and coastal infrastructures.
Hence, creation of new technology is needed by integrating several disciplines such as data management, information systems, knowledge management and others. Full-text technologies are often unsuccessful when applied to this type of queries since they assume the presence of specific keywords in the text; in order to fix this problem, the MAPS project suggests to use different semantic technologies for retrieving the text and data and thus getting much more complying results.
In this scenario the search engine (which is under development) has been designed for operating with structured data – the metadata – and raw data – the associated technical and scientific documentation. The set of metadata associated with marine data is defined in the CDI (Common Data Index) documented standard developed in the framework of the EC project SeaSearch/SeaDataNet. They encode: the types of sizes which have been measured; the measurement tools; the platform which has been employed; the geographic area where measures have been taken; the environmental matrix; the descriptive documentation.
This paper is organised as follows: Section 2 is presenting the related work; Section 3 the methodology adopted and the search engine to be developed; Section 4 shows a case-study and gives examples for using the MAPS search engine.

2. Related work

"Information retrieval (IR) research has been actively driven by the challenging information overload problem and many successful general-purpose commercial search engines. While the popularity of the largest search engines is a confirmation of the success and utility of IR, the identification, representation, and use of the often-complex semantics behind user queries has not yet been fully explored. In this workshop we target methods that exploit semantics in search-related tasks. One of the major obstacles in bridging the gap between IR and Natural Language Processing (NLP) is how to retain the flexibility and precision of working with text at the lexical

level while gaining the greater descriptive precision that NLP provides. We have solicited contributions on automatic analysis of queries and documents in order to encode and exploit information beyond surface-level keywords: named entities, relations, semantic roles, etc" (Workshop on Semantic Search, NAACL 2010).

The term Text Mining (TM) denotes all those semantic technologies that aim to extract structured information from un-structured data (Ramjan et al. 1998). Information Retrieval (IR) groups instead the technologies aiming at retrieving certain information from a document database (Manning, et al. 2009). The currently popular search systems, including those implemented by common search engines, are mainly based on matching of strings: the words of the query are treated as a set of key terms whose presence or absence is sought in previously indexed texts (e.g. Brin, 1998; Kogut et Holmes, 2001)

Semantic search systems are slowly spreading amongst the generalist search engines. They are used in particular in specific domains contexts where the reduced domain terminology makes the adoption of semantic technologies easier. The leading field of these systems is the biotechnology one where the experts need to quickly access to large amounts of information about the literature domain by using specific research related to concepts and semantic relations (Ananiadou et al. 2010, Dai et al. 2010). Even for biomedicine it is important to being able to extract conceptual information about the interaction between the genes and/or proteins; finally, also in the field of environmental sciences significant experiences in this sense have been made.

In general, a preliminary analysis of the texts in question by means of existing technologies relying on lexical and conceptual generic resources it is always necessary for the realisation of a statistical semantic search system in a specific domain. Later on, the results of this analysis – possibly manually corrected - are used for extracting new knowledge through specific terminology that will enrich the lexical resources: in this way these resources will act as a training set for the linguistic analysis systems.

Semantic search systems, like any other search system, may propose systems for search refinement based on metadata or on information previously extracted from the text. One of the most useful aspects of MAPS is the geographic profiling of documents, both on the basis of metadata and of indication of geographical places in the text. In this sense, also systems for the analysis and extraction of named-entities are relevant to the search.

We present here some relevant projects in which semantic search technologies and/or textual analysis have been developed for the extraction and retrieval of information from domain-specific texts:

OpeNER is a project funded by the European Commission under the 7th Framework Programme. Its acronym stands for "Open Polarity Enhanced Name Entity Recognition". The main objective of the project is to provide a set of tools which can be immediately usable by researchers from small and medium enterprises for performing natural language processing activities which are free and easy to integrate into their workflow. Amongst the purposes of OpeNER is to succeed in detecting and disambiguating named-entities as well as performing sentiment analysis and opinion detection on texts, so as to be able to extract from web reviews the feeling and the customers' opinion about a certain resource.

BOOTStrep - Bootstrapping of Ontologies and Terminologies Strategic Research Project - has been funded by the European Commission 6th Framework Programme. The project operated in the biological domain and combined terminology resources and existing databases in a common framework. BOOTStrepp implemented a system of text analysis for the acquisition of new terms, concepts and relations which is a semi-automatic process of ontology-construction in the biological field.

KYOTO (Knowledge Yielding Ontologies for Transition-based Organization), [Co-funded by EU - FP7 ICT Work Programme o2007 under Challenge 4 - Digital libraries and Content] had the goal of providing a system for semantic search in order to allow expert users to model and improve their domain ontology with terms and concepts automatically extracted. The application domain of KYOTO was the environment.

GLOSS - Global Semantic System - The overall objective of the Italian project GLOSS - derived from the KYOTO project but applied to the geographical domain- was focused on building a global system for the creation, use and sharing of documents starting from the current state of the art of the MLIA2MMC technology. The project produced an integrated environment for consultation and sharing of knowledge.

3. Methodology

After having analysed the state of the art, the development of the MAPS project has followed these steps:

1) Collection of needs. This phase focused on the detection of relevant information regarding the content of documents. The information was obtained from interviews and questionnaires to domain users in order to highlight the specifications for designing the engine.
2) Modelling of the Digital Library. This phase was based on the identification of the optimal pipeline according to the users' requirements as collected in Step 1 and on the definition of the requirements regarding the lexical/terminological/ontological resources necessary for the semantic search.
3) Definition and design of the search engine architecture. This third phase was further divided in: a) Detection of specific semantic search interface in order to determine which are the requirements to be met; b) implementation of an effective and performing instrument for document search with respect to the needs of the oceanographic community.
4) Selection of information for the case-study presented in this work.

The design of the semantic engine architecture is based on the integration of already existing linguistic/semantic modules emphasizing the adaptation to the lexical and terminological document bases of the specific marine domain, such as the Italian Marine Semantic Network.

The collaboration between the various partners of the project - ILC-CNR of Pisa and the groups affiliated to the DP2000 and ETTsolutions companies (of La Spezia and Genoa, respectively)- resulted in the elaboration of formats of linguistic and semantic text annotation and of standards for lexical and ontological resources, in order to actually make possible the indexing and querying of documents processed in the database of the Marine project.

3.1. Search Engine

The MAPS Document Base is formed by a set of Italian and English pdf texts and they are the technical and scientific documentation which describes the marine data made available by MAPS. Each text belonging to the catalogue is converted into text format and then processed by a battery of linguistic analysis modules which enrich it with multiple layers of stratified features corresponding to the levels of analysis performed. This expansion is followed by a phase of contraction due to the extraction of semantic content within the text.

With reference to the architecture in Figure 1, the Indexing module collects the extracted content (domain terms, concepts and relationships between them) and it anchors them to the source text in order to facilitate retrieval at a later stage. The Cluster Builder module, instead, organises the extracted semantic information in a domain ontology.

The extraction of information is conducted in several logical steps

- creation of the glossary of terms;
- creation of the network between terms and concepts (ontology taxonomy);
- creation of the profiled text (indexing) belonging to the catalogue of documents.

The terminological glossary consists of a list of automatically extracted terms that – for the English part –will enrich an already existing terminology database: the Oceanographic BODC vocabulary (British Oceanographic Data Center). However, the texts of the catalogue will always be accompanied by the metadata in CDI format (Common Data Index) from which it will be possible to extract some meta-information for creating a small document base for Italian as well.

The name-entities belonging to the glossary may be both simple and complex: the simple terms are those extracted from the NER (Named Entity Recognition) module, that is dates, geographical names, place names, organisations, abbreviations, acronyms, etc. and those terms consisting of a single word, normalised by lemma and filtered by their frequency in documents. The complex terms are noun phrases linguistically normalised which correspond to patterns (such as name-name, name-preposition-name etc.) identified by a sort of mini-grammar. The complex terms thus identified are sorted according to a statistical rank and should be evaluated by applying measures of association (log-likelihood, mutual information, etc.).

The domain ontology is a graph representing a semantic-conceptual network; the network nodes are the terms and concepts of the glossary. The taxonomy is constructed by identifying the hyponymy and hyperonymy relations between terms and this process can be done:

- on a lexical basis - the system is in fact capable of identifying terms with increasing complexity and the sharing of the lexical head is a proof of hyperonymy; for example, if the system identifies as domain terms "boa" and "signal boa" then it determines that "boa" is a hyperonym of "signal boa" and assigns a hyponymy relation in the opposite direction. Similarly, if you find the term “dangerous signal boa" the same reasoning applies to a level further;
- exploiting lexical-semantic resources such as WordNet; in this case, at least for the more generic terms, hyponomy and hyperonymy relations are imported directly from WordNet thus allowing also the extension of the glossary with new concepts.

The introduction of hyperonymy and hyponomy relations concerns the construction of vertical relationships within the ontology. Horizontal relations between terms are partly built by exporting the relationships between entities in Wordnet - if present - but mostly by performing a dependency parsing of texts and identifying the verbs which have a relation with the domain terms.

Finally the indexes allow to efficiently retrieve interesting textual documents when the system is semantically queried on the basis of a specific user’s need; for this purpose the indexing process allows to associate each text of the catalogue to a list of terms and concepts that represent it maintaining the reference to the original text.

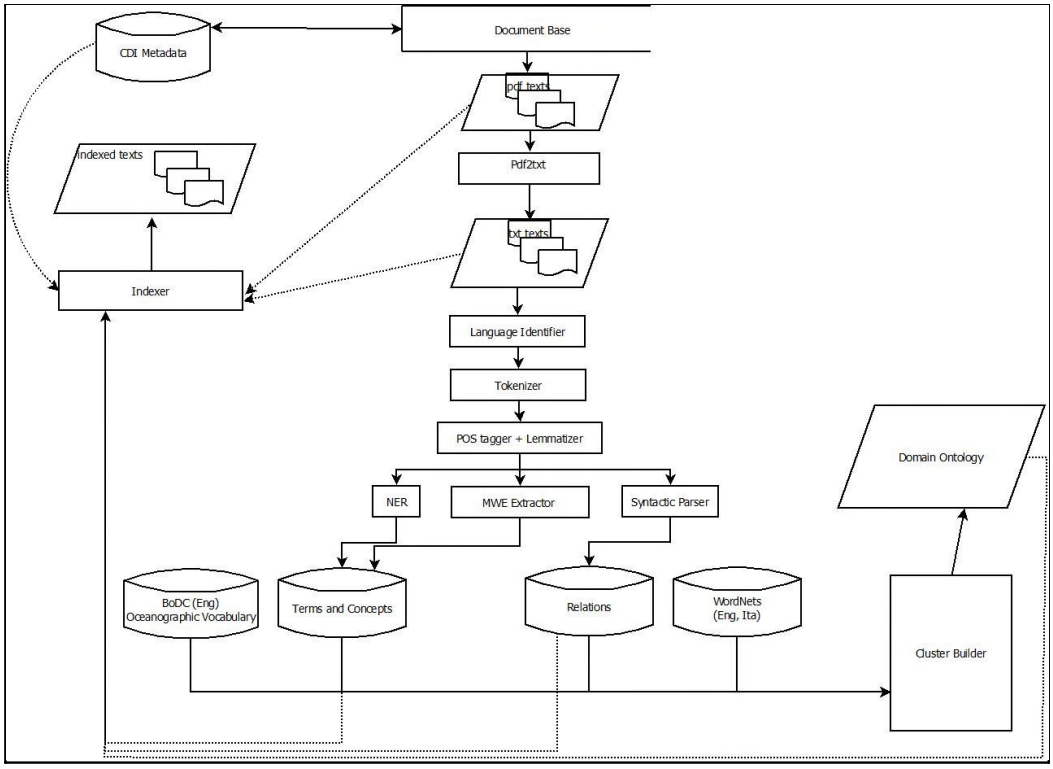


Figure 1: Linguistic Analysis and Indexing

4. A case-study

4.1. Terms

This case-study highlights how the retrieval of grey literature from this specific marine community could be reproduced for other communities as well, thus revealing the great impact that the processing, re-use as well as application of grey data have on societal needs/problems and their answers. The marine application is also making use of common vocabularies that are defined by an international working group called 'SeaVox'.

The illustrative term chosen from the digital library is *Temperature*[1]: this noun has been captured by the semantic engine in the wide variety of contexts used in oceanographic scientific literature. The term *Temperature* is often associated with other nouns, adjectives, acronyms and

[1] For visualising some associated forms of the term *temperature* or some complex terms related to it, please see *Appendix 1*.

abbreviations thus forming complex terms belonging to the domain ontology; these terms in relation with *Temperature* are normalised (lemmatised) in order to be independent from the lexical form: singular/plural, lowercase/uppercase, graphical variants (for example, *sea surface temperature, sea-surface temperature*).
The forms extracted by the oceanographic repository and processed by the semantic engine reveal that the language of this field is mostly constituted by technical terms as well as by acronyms and abbreviations: they might become significant anchors for the information retrieval and the knowledge extraction and this is the reason why the acronym SST (Sea Surface Temperature)[2] has been picked up and the attention has been focused on contexts where it occurs.
It can be noted that in the ontology various concepts gather around the domain term SST: "*temperature*", "*surface temperature*", "*sea surface temperature"* are all terms linked together by hyponymy and hyperonymy relations while "*sea surface temperature"* e "SST" stand for the same concept and are therefore linked by an identity relation.

Figure 2 below shows a few terminological examples extracted from documents where they appear either as a simple term or as belonging to a complex term; the ontological concept has been underlined and is followed by a list of micro-contexts where it is present.

SST	SST Anomaly
• the night time SST	• SST anomalies
• Mediterranean SST	
• satellite SST	SST Retrieval
• a 1/16 degree resolution Black Sea SST	• SST retrieval
• a 1/100 degree resolution Black Sea SST	
• The very low SST	Information SST Level
• The very high SST	• practical information on the SST TAC level
SST measurement	SST Analysis
• in-situ SST measurements	• the OSTIA[3] SST analysis
• valid SST measurements	• SST analysis
• input satellite SST measurements	• the global SST analysis
• buoy SST measurements	
SST Measurement Area	SST_TAC[4]
• input satellite SST measurements in the Arctic area	• SST TAC
SST Satellite Region	Local SST Characteristic
• SST from NOAA[5] satellite for the Australian R region	• local SST characteristics
Real-Time SST	
• near real-time from various satellite SST producers	

Figure 2 – A few terminological examples

[2] *Sea surface temperature*: "Sea surface temperature (SST) is the water temperature close to the ocean's surface. The exact meaning of surface varies according to the measurement method used, but it is between 1 millimetre (0.04 in) and 20 metres (70 ft) below the sea surface" (From Wikipedia).
[3] Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA). http://ghrsst-pp.metoffice.com/pages/latest_analysis/ostia.html
[4] Thematic Assembly Centres
http://en.wikipedia.org/wiki/MyOcean#TAC_.28Thematic_Assembly_Centres.29_Production_Centres.
"The MyOcean project, which is funded by the European Union for a 3-year period (2009-2012), includes the development and operations of a Thematic Assembly Centre for satellite SST products (SST TAC) from (MYOCEAN RDAC PROGRESS REPORT, 2012)..
[5] National Oceanic and Atmospheric Administration (NOAA). http://www.noaa.gov/about-noaa.html

4.2 Oceanographic Semantic Extraction

4.2.1. Digital Library for Grey Documentation

The aim of oceanographic semantic extraction is to retrieve relevant information from digital documents: the innovation of a semantic engine lies in the fact that the process is not just about the retrieval of already known documents by means of a simple term query but rather the retrieval of a population of documents whose existence was unknown.

a) The system answers by showing a screenshot of results ordered according to the following criteria:

- Relevance — Relevance of the document with respect to the "*concept*" which is searched
- Date — Date of publication of the paper
- Source — Providers who share the scientific document
- Matrix — Environmental matrices as defined in the oceanographic field
- Geographic area — Relevance of the geographic area specified in the query
- Clustering — "The process of organizing objects into groups whose members are similar in some way"[6].

≈ The clustering returns as the output the related documents, that is those documents which have in common more than one "concept" described in the ontology.

b) For each document the visualization provides:

- Title, author, source/provider, web address
- Tagging of key terms or concepts
- Summary of the document
- Visualisation of the whole document

c) Advanced search and interaction with the results: the possibility to further refine the search for reducing/increasing the number of results by means of the following partition:

- Typology of formats — Pdf, doc, xls, ppt, etc.
- Typology of material — Video, Audio, etc.
- Language of the document — English, Spanish, French, German, etc.
- Authors' affiliation — Academy, Industry, etc.

The possibility of inserting the number of citations for each document among the criteria of the advanced search is currently undergoing; in this case the engine should be able to connect to any of the existing bibliographic citation systems (such as Google Scholar, Scopus, etc.).

The following is a screenshot of the MAPS semantic search engine:

[6] http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/.

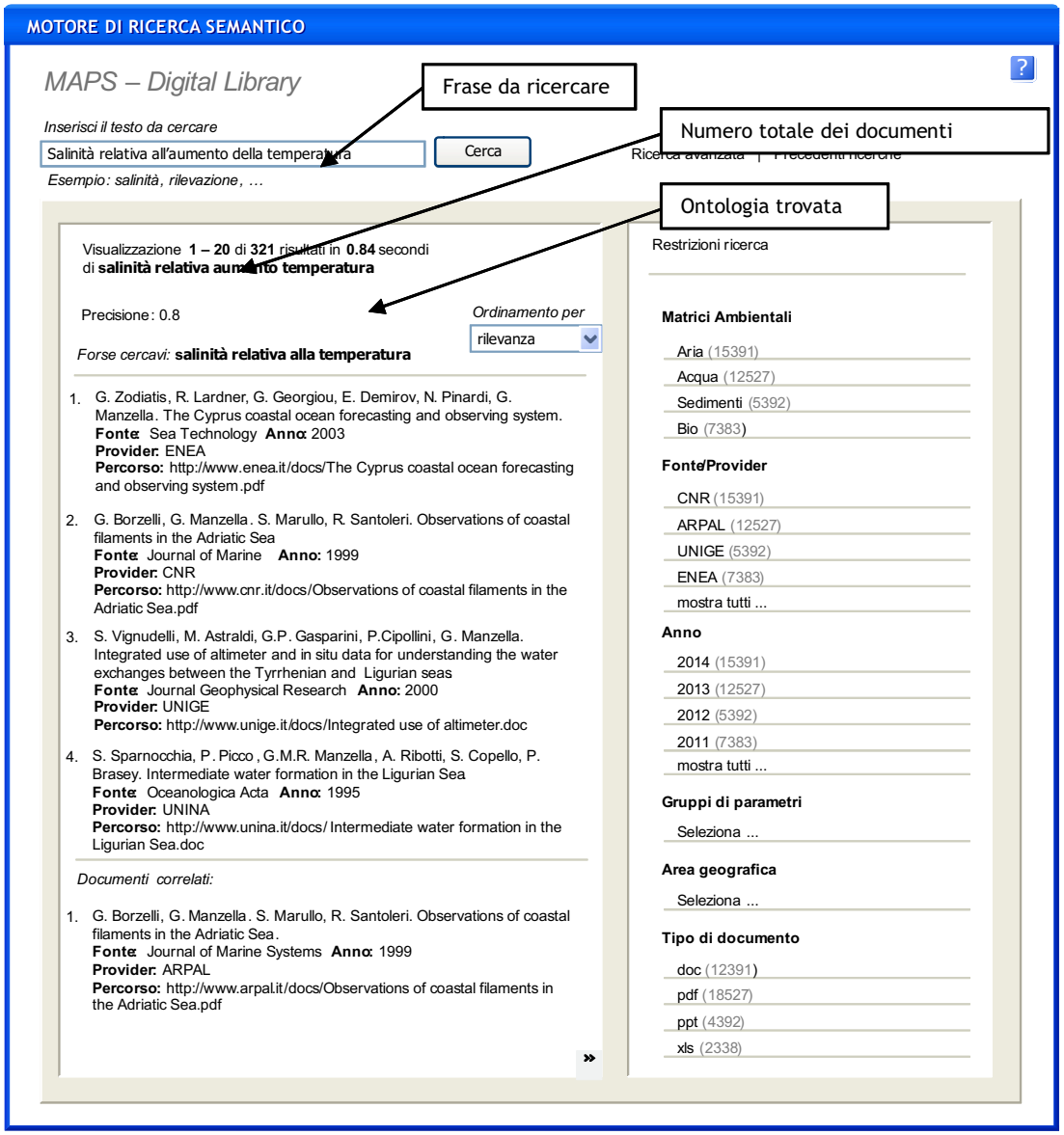


Figure 3 – Search example

5. Conclusions

The MAPS project which has been described in this paper aims ultimately at building a computer platform for supporting a Marine Information and Knowledge System. This integrated MAPS platform offers the advantage to have all content in one place and allows linking different information that exists, thus helping data management activities.

It is well-know that document indexing techniques are not sufficient to satisfy user information needs that go beyond the limits of a simple term matching search. Therefore the MAPS search engine is enriched with semantic technologies aimed at providing an accurate representation of the content of vast repositories of unstructured documents for semantic indexing purposes. Today language technologies make it possible for scientists and developers to produce software applications capable of revealing the semantic properties of textual elements and associating them with conceptual structures. Search functions coupled with semantic-conceptual technologies are able to interpret the underlying search criteria and thus better identify the data and the corresponding documents. This enables an effective and selective access to information even in the presence of significant collections of data.

The proposed solution combines the data available with the technical and scientific documentation thus allowing searching data on the basis of the content of these documents. This makes it possible to select data with greater flexibility, using information that cannot be encoded in predefined data structures. Search functions able to capture technical-scientific

documentation associated to a query constitute an essential feature for the development of advanced services for Operational Oceanography, in sectors such as e.g. civil and military security, safety of navigation and protection of offshore and inshore infrastructures.
While semantic technologies have already been used in other areas for information retrieval with excellent results, their use in oceanography is innovative and responds to information needs that only recently emerged: the interest in the marine environment, so far relegated primarily to the scientific world, is in fact investing new sectors under the actions of the European Community and greater sensitivity of the society for the protection of sea life.

This study has attempted to provide semantic annotations to grey literature documents of the oceanography domain. The initial experiment has revealed that available methods are capable of assisting the process of semantic annotation with promising results. The incorporation of ontologies and knowledge resources in a rule-based information extraction technique promises to enable rich semantic indexing of grey literature documents.
Flexibility is a necessary requirement for any new use that the system intends to embrace while additional efforts are required for further exploitation of the technique and adoption of formal evaluation methods for assessing the performance of the method with measurable criteria.

Acknowledgements
This paper is supported by the Programma Operativo Regionale POR-FESR (2007-2013), Asse 1 Innovazione e Competitività, Bando DLTM Azione 1.2.2 "Ricerca industriale e sviluppo sperimentale a favore delle imprese del Distretto Ligure per le Tecnologie Marine (DLTM) anno 2012. Concessione di agevolazione POS. N° 19. CUP G45C13000940007 MAPS.

References

FRONTINI F., BARTOLINI R., MONACHINI M., PARDELLI G. MAPS. (2014). Stato dell'Arte dei motori semantici . PROGRAMMA OPERATIVO REGIONALE POR-FESR (2007-2013), Asse 1 Innovazione e Competitività, Bando DLTM Azione 1.2.2 "Ricerca industriale e sviluppo sperimentale a favore delle imprese del Distretto Ligure per le Tecnologie Marine (DLTM) anno 2012. CONCESSIONE DI AGEVOLAZIONE POS. N° 19. CUP G45C13000940007 MAPS. Deliverable D1.1,Versione 1.0.

BUSTAFFA F., DE MATTEI M. (2013*). MAPS. Raccolta delle esigenze.* . PROGRAMMA OPERATIVO REGIONALE POR- FESR (2007-2013), Asse 1 Innovazione e Competitività, Bando DLTM Azione 1.2.2 "Ricerca industriale e sviluppo sperimentale a favore delle imprese del Distretto Ligure per le Tecnologie Marine (DLTM) anno 2012. CONCESSIONE DI AGEVOLAZIONE POS. N° 19. CUP G45C13000940007 MAPS. Deliverable D1.2,Versione 1.0.

ANANIADOU, S., PYYSALO, S., TSUJII, J. I., & KELL, D. B. (2010). Event extraction for systems biology by text mining the literature. Trends in Biotechnology, 28(7), 381-390.

DAI, H. J., CHANG, Y. C., TSAI, R. T. H., & HSU, W. L. (2010). New challenges for biological text-mining in the next decade. Journal of Computer Science and Technology, 25(1), 169-179.

FENG D., CALLAN J., HOVY E., PASCA M. (Eds.)(2010). Proceedings of the NAACL HLT 2010 Workshop on Semantic Search [June 5, 2010, Los Angeles, California]. Association for Computational Linguistics (ACL), Stroudsburg, PA. Page iii.<http://aclweb.org/anthology/W10-1200>

MANNING CHRISTOPHER D., RAGHAVAN P., SCHÜTZE H. (2009). An Introduction to Information Retrieval. Online edition, Cambridge UP. <http://nlp.stanford.edu/IR-book/pdf/00front.pdf>

KOGUT P., HOLMES W. (2001). AeroDAML: Applying Information Extraction to Generate DAML Annotations from Web Pages. In *First International Conference on Knowledge Capture (K-CAP 2001). Workshop on Knowledge Markup and Semantic Annotation*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.21.8180&rep=rep1&type=pdf >

BOURIGAULT D., JACQUEMIN C., L'HOMME M.C. (Eds.) (2001). *Recent Advances in Computational Terminology*. John Benjamins Publishing Company, Amsterdam-Philadelphia. Pages viii – xvii.

OAKES MICHAEL P. , PAICE CHRIS D. (2001). Term extraction for automatic abstracting. In Bourigault D., Jacquemin C., L'Homme M.C. (Eds.), in *Recent Advances in Computational Terminology*. John Benjamins Publishing Company, Amsterdam-Philadelphia. Pages 353-370.

BRIN S. (1998). Extracting Patterns and Relations from the World Wide Web. In *Proceedings of the WebDB Workshop at 6th International Conference on Extending Database Technology*, EDBT'98. Pages 172-183.

RAJMAN, M., & BESANÇON, R. (1998). Text mining: natural language techniques and text mining applications. In *Data Mining and Reverse Engineering*. Springer US. Pages 50-64.

The recent Italian regulations about the open-access availability of publicly funded research publications, and the documentation landscape in astrophysics

Monica Marra
Istituto Nazionale di Astrofisica, INAF; Osservatorio Astronomico di Bologna, Italy

Abstract
In October 2013 Italy enacted Law n.112/2013, containing the first national regulations about the open-access availability of publicly-funded research results (publications). The impact of these new regulations with the specific situation of that open-access pioneering discipline which is astrophysics, has been considered. Under a strictly technical point of view, in the light of the new dispositions the open nature of a part of the astrophysical scholarly literature which has been made available online free to the reader during the last twenty years, might be questionable. In astrophysics, most of the journal articles are published by a very small number of scientific societies and organizations. The copyright policies of these major publishing bodies have been collected and analyzed, with regard to the main requirements of the Italian law about open access. Most of these policies are sufficiently liberal for an entirely compliant open access to be provided and scientists would benefit from knowing these details more extensively. Some possible ways to make astrophysicists' scholarly dissemination entirely compliant with the Italian law requirements are considered.

On October 7th, 2013 Italy achieved an important attainment enacting Law n.112/2013, which is the first national law concerning open access.
The law – which in fact pertains to a number of issues related to cultural heritage - had initially been approved in a different formulation (Decreto-Legge n.91 dated August 8th, 2013). During the passage through Parliament, the first stage of the action was subjected to changes, which according to authoritative commentators (De Robbio 2013; Caso 2013) partly deflected the text about open access from the European Union Recommendation on the subject, issued in July 2012.
Law 112, under article 4, decrees:

- OA is mandatory for research outputs/publications derived from research projects which have been financed by at least 50% with public funds;
- if the publication venue publishes at least 2 issues/year (i.e., according to Caso 2013: books are excluded).
- It may be provided a) either on the publisher's site or b) on an institutional or disciplinary repository, by depositing the publication's accepted version within 18 months (STM area) or within 24 months (SSH area) from publication.
- It must be provided at no additional cost.

During the period that shortly preceeded and followed the enactment of Law 112/2013, a series of important initiatives seemed to turn the Italian experts' and concerned people's attention back from the new regulations to the European Recommendation. These initiatives included a) the commitment, signed by the then-Minister for University and Research Massimo Bray, aimed at harmonizing the new law with the European Recommendation (October 3rd, 2013) and consequent actions (i.e., the programme Scientific Independence of young Researchers (SIR) 2014, with open access mandatory clause in the same terms as in Horizon 2020, mentioned below); b) the approval of the 70-billion-euros research financing programme Horizon 2020 at the European Commission, in Autumn-Winter 2013, with European-type open-access clauses; c) the publication of some relevant national statements and recommendations, with a leading role carried out by national academic coordination bodies, also supported by the Presidents of some important national research Institutions, including INAF [1]. These recommendations and statements seem to lean towards the green road to open access and encourage academic and research institutions to issue OA policies and mandates for their research personnel.

In a general perspective, it may be useful to recall that astrophysics is peculiar among other disciplines, with regard to its approach to the dissemination of research outputs. Astrophysics

has actually been so deeply crossed by “the preprint culture” during the last twenty years, that in the opinion of most scientists at international level an open-access policy is considered to be already practised (Harley 2010), the proof being generally indicated in the very large adoption of the practice of posting papers on the ArXiv server in a pre-publication stage.

The literature is not unanimous about the percentage of published astrophysical papers already available in open access mode - which is probably related to the different self-archiving practices adopted by researchers according, typically, to different astrophysical subfields. Actually, archiving of the preprint version of papers occurs as well as archiving of the accepted version does; the (limited) posting of post-print versions as well makes this lively scenario even more rich and motley.

Anyway, although it seems unquestionable that astrophysicists' perception of a widely open-mode available disciplinary literature is well-grounded, we might profitably ask ourselves whether researchers on one side, and regulations about open access on the other, are actually considering the very same object.

In fact the Italian law - as well as other national laws concerning open access (Caso 2013) - considers texts as endpoints of the respective scholarly pathways rather than possible stages of research elaboration and dissemination. Thus, regulations require making free to the reader *the papers' final accepted versions* within a given period - either by publishers on their website, or by authors through systematic deposit on a repository. As unexpected as it may seem, a significant fraction of the impressively long “green OA” road which has spontaneously been taken by astrophysicists during the last twenty years might thus result to be non-compliant with the new law requirements, if considered from a strictly technical point of view.

One of the foremost questions for those technically concerned with open access in general is the following: for open access to be provided on authors' side, do publishers let authors manage their manuscripts freely enough? In particular, and key to our perspective: what do publishers allow authors to do with the final accepted versions of their papers?

The crux is authors' rights as stipulated in copyright agreements with publishers, therefore it is essential to be aware of the main journals' policies with regard to this issue.

In astrophysics, “more than 90% of the original research” is published by a very limited number of renowned international journals (Bertout 2012), almost entirely included in the first and second quartile of Thomson Reuters' Journal Impact Factor (JIF) ranking for Astronomy and astrophysics. I have analysed this set of major journals, which includes ApJ (JIF for 2013: 7/59), ApJL (10/59), ApJSS (2/59), AJ (15/59), MNRAS (11/59), A&A (13/59), PASP (17/59).

[1] On March 21st, 2013 the Conference of Rectors of the Italian Universities (CRUI), together with the Presidents of some important Italian research Institutions, signed the *Position statement on open access to the results to research outputs in Italy* (available in English at http://www.cnr.it/sitocnr/Iservizi/Biblioteche/Position_statement_OA_en.pdf , last visited Nov. 10th, 2014). This statement is envisioned in a European perspective and commits the signers to coordinating initiatives aimed at i) promoting the creation of open repositories, that be interoperable according to international standards; ii) “encouraging researchers to make their research results (data and publications) available in OA journals or Institutional or subject repositories. Research results deposited in Open Access repositories should be made available in their post-print or publisher's version upon publication, and no longer than 12 months after their publication;“ [italics mine]; iii) adopting institutional policies for making deposit mandatory. In January 2014, CRUI issued the Declaration *L’”accesso aperto” alle pubblicazioni scientifiche* (signed jointly with the Italian National University Council (CUN)). The Declaration is critical about the maximum embargo periods stated by the new law and encourages the choice of the “green OA” option (https://www.cun.it/provvedimenti/sessione/140/dichiarazione/dichiarazione-del-15-01-2014 - in Italian, last visited Nov. 10th, 2014).

All of the journals in this set have both a publisher, and a learned society or scientific organisation which is in fact the reference point for the journals' scientific policy and, which is most important from the present point of view, owns the copyright for the journals' publications and may have a copyright policy of its own, different from the publisher's one. The learned societies or organisations involved are the American Astronomical Society (published by the

Institute of Physics Publishing), Royal Astronomical Society (published by Oxford University Press), European Southern Observatory (EDP Sciences), Astronomical Society of the Pacific (University of Chicago Press).

Thanks to the precious collaboration of these learned societies and organisations, the major journals' policies have been individually checked and a comprehensive view of the options which authors are given was gained - as at Spring 2014.

About the so-called “gold road” to open access – which is in fact notoriously questioned among experts as a recommendable choice for institutions -, three versions turned out to be available.

ApJ, ApJL, ApJSS, AJ, PASP and A&A offered a 12-months delayed open access to the entire content of these serials, free to the reader. The time issue here might arise as a critical element in case regulations or mandates (not to mention the authors' wish) asked for a shorter period for the openness to be provided on the publishers' side. The article processing charge (APC) option was not provided – except for A&A -, but the AAS “welcomes discussions” with interested institutions.

The European journals practised a threefold policy. On one hand, A&A and MNRAS provided niches of their content free to the reader upon publication, while making the remaining full-texts openly available after a period which ranged from twelve months (A&A) to the impressive threshold of three years (MNRAS).

On the other hand, they provided research institutions with the APC option, which corresponds to 400 euros per article for A&A and more than five times as much for MNRAS (as at November 2013).

Obviously, an expenditure issue may arise in some of these cases.

For as much as the “green road” is concerned, ApJ, ApJL, ApJSS, AJ and PASP let authors post the accepted versions of their articles freely on repositories, “whenever they like” for the AAS and also immediately - provided there is an institutional mandate requiring deposit in fewer than twelve months - in the case of PASP.

“Astronomy and Astrophysics” let authors self-archive the accepted version of their papers before publication, but asked researchers to substitute this version with the published one, once the latter is available.

MNRAS let authors self-archive exclusively a PDF version of the final published article, upon publication (as at November 2013).

Bearing in mind all of these copyright policies and the law requirements; considering the citation advantage for papers disseminated in a pre-publication stage; and being notorious that astrophysicists' scholarly routine is in most cases already inclusive of a timely, pre-publication self-archiving of papers on a repository; we might conclude by endorsing the Green Road to open access as a first choice option, thus supporting the recommendation issued by the Italian Conference of Rectors (CRUI), which was signed by the Italian National Institute for Astrophysics on October 4th, 2013 [2].

[2] CRUI and some relevant Italian public research Institutions recommend “ [...] the adoption of policies and institutional regulations which make deposit in their institutional repositories mandatory; in case the latter don't exist, publications and data should be deposited in other institutions' repositories or in disciplinary OA repositories”.

I

In case the economics of these processes are a matter of concern, the peculiar modes of dissemination practised by astrophysicists, which have hitherto been characterized by the “productive coexistence” (Henneken et al, 2007) of established journals on one side, and very wide networks of versioned, free online availability on the other side, might preserve the sustainability of the learned societies' activity, or in some cases (MNRAS) hopefully make it more generally affordable to research institutions.

A final, additional consideration deserves attention. In spite of their central importance within astrophysics, the refereed journal articles seem in fact to represent a minority share of the

discipline's research output (less than 40% according to the sample study by Marra, 2013). In the spirit of making open access progressively more common practice, the use of a Current Research Information System (CRIS) as an official tool for compliance with the OA provisions might open the way to a similar treatment for the other research and technology output types which are or might be hosted by this kind of tool, and which cooperate significantly in the shaping of the astrophysical documentation landscape.

References

Bertout, Charles (2012), *A brief overview of the publication process in astronomy*, in *id*., *A guide to effective publishing in astronomy*, Les Ulis, EDP Sciences, 2012, 13-18.

Caso, Roberto (2013), *La legge italiana sull'accesso aperto agli articoli scientifici: l'inizio di un percorso normativo*, in ROARS. Return On Academic ReSearch (blog), 6.11.2013, http://www.roars.it/online/la-legge-italiana-sullaccesso-aperto-agli-articoli-scientifici-linizio-di-unpercorso- normativo/

De Robbio, Antonella (2013) *L'Open Access è legge anche in Italia*, in *Il BO. Il giornale dell'Universita' degli Studi di Padova*, 25.10.2013, http://www.unipd.it/ilbo/content/l%E2%80%99open-access-e-legge-anche-italia

Harley, Diane et al. (2010), *Astrophysics case study*, in *id., Assessing the future landscape of scholarly communication: an exploration of faculty values and needs in seven disciplines*, Berkeley, Center for Studies in Higher Education, 2010, 138-206.

Henneken, Edwin et al. (2007), *E-prints and journal articles in astronomy: a productive coexistence*, "Learned Publishing" 20(1), 2007, 16-22.

Marra, Monica (2013), *Astrophysics as an excellent object for bibliometric analysis, myth or reality? A sample study comparing bibliographic coverage in NASA/ADS, Scopus, Web of Science and in the institutional database CRIS-INAF* (not published), DOI: 10.13140/2.1.1493.8561
https://www.researchgate.net/publication/261911739_Astrophysics_as_an_excellent_object_for_bibliometric_analysis_myth_or_reality_A_sample_study_comparing_bibliographic_coverage_in_NASAADS_Scopus_Web_of_Science_and_in_the_institutional_database_CRIS-INAF?ev=prf_pub

Data: Is it Grey, Maligned or Malignant?

Julia M. Gelfand and Daniel C. Tsang
University of California, Irvine, United States

Abstract:
Cancers, growths, past events, social issues, conditions, and trends are each proverbially described as on a spectrum from maligned to malignant and scientists, physicians, journalists, commentators, politicians and other specialists offer opinions and commentary on what frames the answer to this question of the title. This paper explores not just the color and tone of data, but attempts to resolve what characterizes whether data is maligned or malignant. Hues of greyness distinguish the perils of failing to share, publish nor make accessible research data and the contemporary consequences to scholarship and open access are critical objectives in today's information arcade. Access to data is determined by those who can afford it, discover and know about it, and can thus manipulate it. Grey literature can take the offensive approach to further the role of data, and promote it to advance the common good, contribute to social responsibility and human actions. Data, while increasingly ubiquitous and abundant is the driver of evidence-based foundations, and the link to academic credibility, communication, discourse, dialogue, and the platform for greater open access. Grey data, possessing some of the attributes of grey literature, difficult to identify, acquire and access, when endangered or threatened, not archived or preserved, requiring methods to organize, sort and stratify, forces nontraditional publishing to pursue data publication to enhance perpetual access and new interpretations for its utility in future learning and research applications. We know that there is a somewhat elevated likelihood that open data policies lead to more widespread knowledge and information sharing, greater self-confidence among information providers and scholars alike, but we know less of whether these patterns have any short or long term benefits or disadvantages for individuals or society and of the factors that moderate and mediate these effects. In the meantime, the new reality is that data is central to the work of science, social sciences and basic human conditions of health and wellbeing and data policies mostly proceed from a grey containment to this new reality. The argument that as libraries become active publishers by digitizing content, creating new content, supporting researchers by addressing new domains and formats, that other interpretations of grey data and data more generally are increasingly plausible and that further research on the factors moderating and mediating the effects of data management is needed. This paper explores the continuum for data from maligned to malignant and anticipates data approaching the benign stage emphasizing new hues of grey and open access.

Definitions of Grey Literature & Grey Data

There is much to reconcile – the analogies versus disconnects between Grey Literature and grey data and the relationships between maligned and malignant and how those terms can apply or describe the expanding sphere of grey data today. The grey literature movement with origins now nearly four decades old documents how parallels between original grey sourcing impacts new components of scholarship such as data in the contemporary research setting. From the 1970s when technical and government report literature was the domain of grey literature championed by Charles Auger in his early compilations[1] or a more expansive definition, coined in 1997 that became known as the "Luxembourg definition," which included, "that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers."[2] An additional clause was added in 2004 noting that grey literature "is not controlled by commercial publishers, or "where publishing is to the primary activity of the producing body."[3]

Fast forward to 2010 when this conference met in Prague and there was a call for some revisions to the definition that reflected more online, electronic or born digital publishing. It was proposed then that four attributes should now qualify the basic definition:

- Documents character of grey literature – more multidisciplinary in content and form
- Includes legal nature of works of the mind – protecting intellectual property
- Reinforces quality level for peer review, validation
- Links to intermediation bridging collection status to readers or users

And the new "Prague Definition" for grey literature was defined: "Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers, ie, where publishing is not the primary activity of the producing body."

This definition has been challenged as still being too narrow considering new forms and practices of scholarship and research underway and the methods of publishing now widely available. The emphasis on different forms of publishing and the added dissemination streams of content reaching people in new ways; the constant influx of new technologies; incorporation of new media; containing raw research including data, interviews, previews; new forms of single and combined authorship; and the added concerns for preservation, stewardship and perpetual access regardless of future revisions; are all concerns for a current definition of what constitutes being grey or called "grey literature."

Data for the purposes of this paper is defined as research data and we have embraced the standard OMB definition, "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings."[4] Data can be as effusive as grey literature, reflect disciplinary intersections as well as stand on its own and be as varied as spreadsheets, laboratory or field notebooks, codebooks, protein or genetic sequencing, spectra, test plotting, samples, specimens, digital objects, database elements, models, algorithms, formulas, simulations, methodologies, workflows, operating procedures, protocols, and the like. What research data does not include are trade secrets, commercial information, information protected under law, and personnel or medical information that would clearly constitute a clearly unwarranted invasion of privacy. Balancing privacy and data usage will remain one of the biggest challenges in working with data and will contribute to the ongoing "greyness" associated with data.

Data Science as a discipline is also becoming a programmatic area that crosses mathematics, statistics and economics and spans across all disciplines. The definition at our own institution includes how it "may encompass the full spectrum of theories and methods that use data to understand and make predictions about the world around us." This includes "fundamental research on statistical methods, prediction algorithms, data management techniques and policy issue; as well as a broad range of domain-specific data-driven research problems in the sciences, engineering, humanities, education, medicine and business/management."[5] It is a broad range of "theories, algorithms, methodologies and tools that allow us to use data to better understand and make predictions about the world around us."[6]

Some examples of common elements that bridge grey literature with grey data include:
correspondence, project files, grant and ethics applications, technical reports, research reports, signed consent forms.[7]

Maligned or Malignant?

Why, you may inquire do we pose the spectrum of maligned or malignant to define our thesis? Maligned is used in the sense of speaking ill of someone or something, being injurious, slanderous or defaming by causing one to conclude very differently about given circumstances. An example may be when shown to be evil in disposition, nature or intent in order to conclude different outcomes. Malignant, known for its pathological or cellular definition of tending to produce a bad outcome such as death from cancer and can be highly invasive, dangerous, cause harmful influence or effect suggests something out of a researcher's control but will have to be explained due to some other force or context. Thus our imagery shows the extremes of confusion to an example of the recent anthrax episodes that generated fear and uncertainty about the potential exposure and outcomes.

We were influenced by the psychologist, Professor K. Alison Clarke-Stewart who framed some important questions about social issues such as infant day care as "maligned or malignant?"[8] She

assessed what kinds of available data perhaps best suggest there could be separation anxiety among infants from their mothers who returned to work during their early infancy and what risks these infants would have for future emotional insecurity and social maladjustment when placed in daycare. She also asks what the empirical evidence is concerning the effects of infant day care – "it is truly bad for babies or it is undeservedly maligned,"[9] and goes on to examine what evidence there is to support either supposition.

The conclusions of social science and other forms of scientific research reveal that there is much to learn about the effects of something being studied or the outcomes. We can conclude that perhaps there are elevated or declining likelihoods of something happening, we can establish patterns, review longevity data, and understand that different forces contribute to maturation, change and outcomes. We know about situational evidence or trends in work environments, lifestyles that can influence outcomes without observing or studying those contributing factors, such as climatic or environmental occurrences, catastrophic or unexpected events, political interpretations or activities, market surges or declines, and thus the issue today is not whether data is sufficient for the research being undertaken but what else can we learn from the entire research experience and landscape that influences what we observe and learn. That begins with the research question, how the research is conducted and whether sufficient data is collected to offer any conclusions, definitive or not.

Perceptions of data by different communities

Data used to be the stepchild of the research publication process – considered as not worthy of including with a publication, too complex to describe adequately, or too bulky to attach and beyond the needs of most readers. Funding agencies until recently never gave much thought about how to manage research data, even data generated with taxpayer funds, so data were left to rot literally (as computer tapes or disks deteriorate over time), without any thought of preservation and retention of the data underlying scientific research. Like grey literature, research data can be termed as largely fitting the concept of grey data – unfit to see the light of day. They tended to be left on floppy disks or other removable storage devices, often unintelligible to anyone but the original data compiler and over time were unable to be deciphered. Unless the data are collected as part of a well-funded large-scale study, individual datasets are often forgotten after the research is finished and a paper produced or publication issued. Although the scientific ethos mandated reproducibility, research data was more often than not left in filing cabinets or on computer drives inaccessible to others.

That has begun to change. With the advent of concepts like Data Paper[10] and the development of mandates from US federal agencies including the National Institutes of Health (NIH) and the National Science Foundation (NSF) to share and deposit research data for public access, one suspects that data has suddenly been elevated and given a more prestigious status. The concept of data paper is to create a searchable metadata document where instead of focusing on a journal article where hypotheses and conclusions are central, the data becomes the primary emphasis. Increasing examples of this are known. This has been done by the environmental sciences community that studies biodiversity, and these scholars achieve recognition like in other forms of scholarly and creative outputs, by increasing the visibility, usability and credibility of the data resources now cited and published.

When Megan Smith, Chief Technology Officer of the United States responded in an interview when asked how her office is working to make large sets of government data public, "Scientists and universities and the general public can do extraordinary things with it. It could be weather or climate data; it might be data from the Department of the Interior or NASA or water data. Whole industries are being built from things that taxpayers have helped the government know."[11]

Still, efforts to develop new ways of managing research data while encouraging researchers to share their data have been stymied at the individual level. Researchers may be unwilling to share their data, or see little utility in making it available to a public likely to be uninformed about how to interpret the data. Efforts to teach individual researchers about research data management

are time consuming and in the main not supported by funding agencies. Funding for research data management is more likely to become more available to institutions than to individual researchers or research projects. Also, the mandates encourage repurposing data may become a more central practice than constantly funding variations of the same research question. Data has demonstrated that it is more than "nice to have," or being relegated to second class citizenry, for when data is misunderstood it changes the tenor or disposition in fundamental and profound ways.

Risks & Conditions of Data: Unevenness, authoritativeness, misinterpretation

Data is often easily misinterpreted, especially if research results do not contain enough information to place the research in context. How the data was collected, the sample size versus the universe, whether the data was adequately described, and who funded the research are issues any outsider might want to have answered.

Metadata is often inadequate and inconsistent, filenames are subject to private whim, and documentation of steps in the research process are often lacking. Researchers thus can make bold and extravagant statements about research findings – unfortunately not backed up by their data upon closer perusal. A study may not be in fact generalizable but that may not deter the mass media to tack on to the latest research and proclaim a scientific "breakthrough." Traditionally research subjects in some fields were limited to who were readily available, say, college undergraduates of a certain ethnicity and demographic. Whether such findings are in fact generalizable to a larger population, another setting, or another time period is dubious.

Repurposing the research frontier

With the last few decades demonstrating that an increasing amount of research is done collaboratively with different specialists sharing how to examine and study a problem, bringing their specialized skills and objectiveness to the frontlines, and having the desire to share research findings with other likeminded citizenry, several conclusions are evident. They include the need to share research data widely. This can be from the vantage point of international or global sharing with the focus on making sure colleagues and inhabitants in developing countries who are already challenged by not having as dependable or consistently good access to information, get it.[12] It also serves the ongoing development of ICTs, statistical analysis and the intensity of interdisciplinarity across many subject areas and the entire academic landscape.

The directive of working or conducting research in a more open and networked environment, now increasingly commonplace suggests that the cyberinfrastructure that we come to rely upon is more complex than we perhaps anticipated as it was launched. With the unknowns came opportunities and the role of the Internet over the last two decades has promoted connectivity, exchange, collaboration, and reduced barriers and thus redefined new indicators for competitive intelligence and for intellectual competitiveness, a sense of evidence-based models and the ongoing need for best-practices. Supercomputing now available in a handheld device offers new metrics and ease of computation that can be achieved more economically and efficiently. Being committed to the tenets of sustainability, whether it be economic, social or environmental[13] reinforces how the research process ensures integrity, accessibility and stewardship of research data in the digital age.

Data is often delegated to the cloud. Out of sight but not out of mind, data has found a parking spot in the cloud where costs are less and options more plentiful in terms of what to park.

Role of Libraries: Finding & supporting data

Library catalogs, finding aids and most recently the discovery systems that consolidate access to all content via this new breed of discovery tools support data in new and novel ways. By engaging with social networks, media, and other innovative technologies such as QR Codes, discovery systems from a user, contextual and content owner or publisher perspective will impact a greater reliance on data.

The Internet Manifesto 2014[14] widely adopted by the library community and its publishing partners ensures equitable access to the Internet and its services in support of freedom of access to information and freedom of expression and data related content is no exception to this. As the current chair of IFLA's Committee on Freedom of Access to Information and Freedom of Expression, (FAIFE) writes about this update,
"The world has changed significantly since 2002 both physically and digitally, and we now have a greater experience and understanding of the role of the Internet and digital resources in our services, and in developing connected societies where individuals have the skills that they need to exploit the opportunities that technologies can bring. We also have a greater understanding of the threats that can be posed through the Internet including the impact on human rights of inappropriate monitoring and surveillance, and from criminal activity….reflects this experience and reinforces the vital role of library and information services in ensuring equitable access to the Internet and its services in support of freedom of access to information and freedom of expression."[15]

Libraries, even experiencing these trends remain insufficiently staffed and ill-equipped to handle research data management, but they welcome the opportunity to manage data much like they have taken on different elements of library holdings. Managing research data is not the same as oversight for digitizing printed materials or archiving and preserving digital objects. Data needs to be curated, backed up but also restructured as formats and platforms change. In other words data curation is not just dumping data in a repository, providing access, telling users where it is and considering that sufficient. The level of data curation will vary depending on the resources, skill-sets and time a library or institution can devote to it.

Typically a library or data repository receives a dataset at the end of a research life cycle, often decades after – thus posing often intractable challenges to the data archivist.
Efforts now are being devoted in some institutions to involve researchers at the beginning or during the research life cycle in best practices in research data management. But such efforts are labor intensive, varies by discipline, and often unfunded.

Data Professionalization
With the emphasis of this paper positioned on the role of libraries and librarians handling research data, we should not discount how the commercial world is also responding to the need to create more reliable data management services. The advertising industry is partnering with the computer industry to explore how it "turns Big Data into new ideas." Ogilvy & Mather, among the largest agencies with global revenue, has a new idea about stepping up a longstanding commitment to data-driven decision-making by forming a unit named OgilvyAmp where "Amp" is short for "amplify" and is being directed by a leader with the new job title of global chief data officer. The previous emphasis on information evangelists and such gurus has led to this latest series of tasks that include data strategy and planning, analytics services and data management continuing the way innovative companies such as Ogilvy Mather has amplified the creative process when advertising is developed by paying attention to data with sayings like "Advertising people who ignore research are as dangerous as generals who ignore decodes of enemy signals." The goal of companies such as Ogilvy is to "remove data as a distraction and position it as a tool in the creative palette." Descriptions note, "As data becomes more complex, we want to move data up in the process I support of the creative…You need someone to translate stories out of the data instead of just giving the creative a dashboard." "Creative departments should embrace data as part of their raw material, instead of seeing it as taking their power away." "Data doesn't replace creativity…but great creativity has lasting power that data may not have predicted."[16]

In other innovative business and management applications, data has value if it contributes to efficient and cost-effective savings. Data gains tractions if it has positive outcomes for:

- Reduces building & investment costs
- Allows for better and improved cash flow during projects
- Encourages new opportunity costs – using new apps to interpret results of search or data analysis
- Promotes a more green and sustainable environmental landscape[17]

Data mastery is highly associated with a digital vision. As companies and organizations create their digital vision, they are experiencing a digital transformation, although a compelling vision of a digital future still remains unclear. Technology can remove obstacles and extend capabilities but to enhance users or customers' experiences, streamline operations and transform operational methods and business models is a never ending process.[18]

Back to the library

Today we see positions posted in nearly every research library around the world trying to recruit for professional librarians who can offer appropriate and relevant services to be delivered by a team of research data curation librarians. Sometimes called eScience or eResearch Librarians, Data Librarians or Digital Initiatives Librarians, the work that these professionals now conduct reflects the scholarly communication environment we find very commonplace, where we want to engage users as authors, and develop collections and services that directly support teaching, learning and research missions of the institutions that are trying to re-imagine the contemporary research library.[19] These positions are commanding a lot of attention from library users seeking directions in where to publish, how to publish, how to provide accompanying research data and how to share and repurpose data. Not easy to fill, there remains insufficient competition for all the slots that are being created. Related is the province of statisticians and indicators of professional forecasting does not suggest that there will be an adequate supply of them either for the needs in the marketplace.

In addition, new data intensive environments rely upon the work of semantics, digital humanities, web technologies, human computer interaction, data-mining, information retrieval, communication, strong web programming skills to forecast for tomorrow's new discoveries. As Steve Lohr writes, we still must be mindful that this popular new trade contains far too much handcrafted work or "data wrangling," "data munging" or "data janitorial work." It has been estimated that between "50-80% of data scientists' time" is devoted to "collecting and preparing unruly digital data, before it can be explored for useful nuggets."[20]

Repositories

Many libraries increasingly have institutional repositories and digital commons to contain the scholarship from their affiliates. These repositories are in addition to the thematic or subject repositories that have evolved where one stop deposit and consultation is increasingly the norm to stay informed about the latest contributions to a field. Examples of this include ArXiv, with nearly a million ePrints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics (http://arxiv.org/), the Social Sciences Research Network (http://www.ssrn.com/en/) that supports a score of subject areas. There are also major incentives to not disenfranchise non-recognized or non-distinguished members of the academy but to encourage undergraduates and the student population to also deposit their creative outputs. Examples[21] of journals containing student work, independent projects and group work are increasingly gaining recognition as seen by the detailed work at Illinois Wesleyan University, a small independent liberal arts institution. Such efforts are examined in the increasingly common efforts to promote open data, open education and just reduce the barriers for access. A recent emphasis that addresses these issues is the intersections between scholarly communication and information literacy that librarians have been embracing for the past few years where new library services have been developed to offer supplemental instruction about the value of each of these significant roles.[22]

Just as we continue to explore grey literature, grey data, open data, open science, open education we celebrate ten years as plans are underway for the Tenth International Conference on Open Repositories to be hosted cooperatively by the Coalition of Networked Information (CNI) and several institutions in June 2015. That proposed agenda is a call for many topics to be explored including how repositories can support the needs of research data via data registries, storage, curation lifecycle management, management and digital preservation tools. In addition, the CNI is also tracking the establishment of digital scholarship centers or labs in universities and colleges and hosted a program in April 2014 that offered participants a way to examine the

intersections of Scholarly Communication, Visualization and Digital Scholarship.[23] The data element was very visible in this work environment.

How data is structured

At a recent forum on our campus highlighting the new Data Science Initiative[24] we learned how distinctions between statistics and computing have merged and blurred. Related to this may be the parallels of machine learning to the operations of brain functioning. IBM has Watson and it has been shown that learning has similarities between the human and a programmable robot. The same deep learning problems can be applied across the sciences to physics, chemistry, biology and related disciplines. Depending on computer science, programming, machine learning, statistical analysis, varied technologies, one can integrate applications with understandings so that they can be ready for analysis.

Impacts of Open Access

The Open Access movement has contributed significantly to perceptions, expectations and directions of open data. Publishing opportunities and scholarly communication directives illustrate many examples of how government and the private sector have independently and collaboratively produced new information products, resources and methods of handling data. The pricing of free content does not always come without other associated costs and as soon after the launch of the Educational Resources Information Center (ERIC) by the US Department of Education in 1966 we have seen a very broad range of products and methods of information production.

Highlights include the following launches as noted on Peter Suber's Open Access Timeline:

- Agricola database in 1970 from the National Agriculture Library
- USENET created in 1979
- Text Encoding Initiative (TEI) in 1987
- Psycoloquy among first free online journal that became peer-reviewed in 1989
- First web page was debuted by Tim Berners-Lee in 1990
- Preprint Archives, Mathematical Physics and arXiv launched in 1991
- GenBank launched by the National Center for Biotechnology Information (NCBI) in 1992
- CERN launched its preprint server in 1993
- Human Genome Project, Social Science Research Network (SSRN) and NASA Technical Report Server went live in 1994
- Scientific Electronic Library Online (SciELO) went online in 1997
- Wikipedia was born in 2001
- Budapest Open Access Initiative (BOAI) becomes reality in 2002
- Several scholarly communication initiatives launched in 2002
- Public Library of Science (PLoS) obtains funding to start tow open access journals in 2002
- Google began to digitize and index millions of public domain and copyrighted books from five major libraries in 2004
- From 2004 on, aggressive activity on a global scale took place exploring how to produce open access content and share worldwide – from governments, academic institutions, non-profits, philanthropic foundations, and hosts of individuals with journal declarations being among the most common examples of new information resources[25]

Examples of Data

One flavor of data does not apply across everything. Some will be greyer than others but making generalizations remain problematic. One should be careful and cautious when defining data and its properties. Some examples and current experiences follow:

Survey Data

Common elements include but are not restricted to consumer behavior, trends and, demographics. In the social sciences, survey data is typically structured as a rectangular alphanumeric file, arranged by respondent responses to multiple-choice questions. Areas where such data are relied upon include demographic, public opinion, political participation and

consumer behavior research. Given the current importance of social media in social movements etc., researchers are likely to become less interested in structured survey data and will want ways to analyze data that are structured differently, e.g. as tweets. In other words data mining will become a necessary core skill.

Big Data

Getting the basics down for understanding Big Data is easily done by reviewing Cathy O'Neil and Rachel Schutt's book, *Doing Data Science*.[26]

The report, *Big Data: Seizing Opportunities, Preserving Values* issued by the White House this year and currently under review[27] suggests the importance placed on big data due to the increasingly available sensors, cameras and geospatial technologies that can track global movements. The purpose of this report was to determine how big data will transform the way we live and work and alter the relationships between government, citizens, businesses and consumers. The perception is that big data technologies will be transformative in every sphere of life. Big data according to some experts is "fundamentally reshaping how Americans and people around the world live, work and communicate. It is enabling important discoveries and innovations in public safety, health care, medicine, education, energy use, agriculture and a host of other areas. But big data technologies also raise challenging questions about how best to protect privacy and other values in a world where data collection will be increasingly ubiquitous, multidimensional and permanent."[28] As exciting as knowledge discovery is, and how intensely the Internet has changed how we live our lives, big data as it is defined does not come without challenges and concerns.

The definition of big data is "The capability to manage a huge volume of disparate data, at the right speed and within the right time frame, to allow real-time analysis and reaction." Big data is typically broken down by three characteristics, including volume (how much data), velocity (how fast that data is processed), and variety (the various types of data).[29] Data virtualization and data warehousing is critical for businesses as well as for research practices.

What the US Federal Government has learned so far is that there is concern with data practices.

- Big data tools can alter the balance of power between government and citizen
- Big data tools can reveal intimate personal details
- Big data tools could lead to discriminating outcomes[30]

The policy recommendations made so far include:

- Advance the Consumer Privacy Bill of Rights
- Pass National Data Breach Legislation
- Extend Privacy Protections to non-US Persons
- Ensure Data Collected on Students in School is used for Educational Purposes
- Expand Technical Expertise
- Amend the Electronic Communications Privacy Act[31]

It is easy to agree with Emma Uprichard that we are all consumed by the "challenges of big data." We are led to believe that big data "brings new hope to big social problems and social policy that big data will help us deal with crime and terrorism, intervene with social problems and social policy and may be cheaper to use than organizing large-scale official surveys."[32]

From this we can conclude that technologies are driving the potential for current practices that may accurately or inaccurately represent what we value most. In the spheres of science, technology, and medicine, some ongoing concerns include, data formats, the ambiguity of human language, the need to repeat applications in data projects.

Examples of tools that make data less grey include:

- Spreadsheets – Excel has been around a long while and offers methods to create charts and calculations to allow for models

- ClearStory Data[33] – software that recognizes many data sources, pulls them together and presents the results visually
- Statista[34] – commercial sourcing of data that gives public & private data for eSearchers information in multiple options of delivery offering infographics
- Trifacta[35] – uses machine learning technology to find, present, and suggest types of data that might be useful to see and explore
- Paxata[36] – automates data preparation for analysis
- MapReduce[37] – a software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers
- Data Conversion Laboratory (DCL)[38] – been in the marketplace since 1981; uses optical character recognition (OCR) to integrate data from multiple sources such as scanning, MS documents,

Each of these can be applied to social science data, medical, environmental, science, and interdisciplinary contexts that blend and create new understandings and lend to new knowledge generation. Data and digital conversion, storage, archiving, description and analysis are all the more possible due to "automation and technology that lets us transform things without needing to build things anew every time."[39]

Credibility of Data

"Big data may mean more information, but it also means more false information," and even when the information may not be false the problem is "that the needle comes in an increasingly larger haystack."[40] That is the problem most researchers and scholars face.

Contemporary studies of mindfulness and how it contributes to creativity and innovation have been in both the scholarly and popular media. "There are some who believe the increasing power of Big Data (using powerful computers to sift through and find patterns in massive amounts of information) is going to rival the human consciousness at some point. But there's also growing skepticism about how effective Big Data is at solving problems."[41]

All data are not equal – "experts" may "trust" certain data over others. Source credibility is one big factor – who is compiling the data and is that institution's work reliable? Are its methods transparent? And is there enough documentation to allow a subsequent user to understand how the original data was collected, arranged, described, massaged, and cleaned. This has led to the concept of apply a data seal of approval to data archives – for example, ICPSR, the major social science data repository at the University of Michigan, has obtained a "Data Seal of Approval" for its work.

Researchers are concerned about relationships and conclusions that their findings and data suggests. Spurious correlation, defined by Pearson in 1897 describes the correlation between ratios of absolute measurements that arises as a consequence of using ratios rather than because of any actual correlations between measurements[42] and is one of the main motives for the field of compositional data analysis which deals with the analysis of variable that carry only relative information, such as proportions, percentages and parts-per-million.[43] Spurious data, relationships and correlations may offer respite and humor but the seriousness of when they enter mainstream media or are introduced erroneously is problematic. These examples of relationships of total revenue generated by computer arcades correlating with the number of computer science doctorates awarded provides mixed up interpretations as the computer gaming industry matures. Another example is the number of oil imports correlating with consumption of high fructose corn syrup.[44] Pearson's definition is obviously not to be confused with misconceptions about correlation and causality.

Data Publishing & Publications

New information products issued by a variety of sources globally support individual researchers and functions of work conducted in different sectors. Specific resources developed by the academic community with which we are most familiar include the Data Management Plan

created by the California Digital Library (CDL), and other widely respected efforts generated at Purdue, in the UK and at many institutions.[45]

The University of California has also created DASH, a data sharing platform for researchers at the 10 different campuses.[46] These resources are open source and can be used by institutional adoption at other institutions. Each of them contributes in unique ways to data sharing and ultimately to new knowledge generation.

Funding agencies requirements vary regarding the retention periods for data generated from sponsored research, depending on discipline, and form of data. Data longevity thus may vary by different indicators. This implies that data can be "weeded" – a heresy to some data archivists whose archives had missions to preserve and retain data in perpetuity. Before deciding to weed data, attention needs to be given to whether the data can be repurposed to meet current research needs, and whether format migration is feasible. The retention vs de-selection debate remains controversial and problematic, with tools and economic consequences allowing for a longer lifespan and less expensive storage of data with ease of providing better metadata describing the data.

Looking Forward: The next horizon for grey data

Recent research and scholarship has shown a large wave of interest in data management in this accelerating digital age. We must continue to be vigilant about the "validity of research data, standards that do not keep pace with the high rate of innovation; restrictions on data sharing that reduce the ability of research to verify results and build on previous research; and huge increases in the amount of data being generated, creating severe challenges in preserving that data for long term use."[47]

Reflecting on the Internet Manifesto, the current trajectory and the landscape we think data, especially grey data instills, we can't forget the diverse nature of library work and the opportunities we now have with diverse technologies and new skills individuals must develop and realize that it too, will continue to be part of an evolution of services and resources in libraries. As Steven Bell writes, "But what if, in addition to other core values instilled in LIS programs such as protecting patron privacy and defending intellectual freedom, LIS students learned and practiced methods for tackling tough problems and developing thoughtful solutions – in any situation. Academic libraries need LIS graduates who can assess situations and resources, identify the source and nature of a problem, and then craft an appropriate solution. In other words, educate [the next generation] to identify, frame, think through and solve problems the way designers do."[48] Perhaps we will have strategists, technologists, designers and curators in the role of librarians to manage data.

Conclusions: Maligned or Malignant?

If scholars think that data is maligned it may be because of this "liberal notion" we possess regarding how evidence-based decision-making is superior to one without evidence. The specialty field of logistics and decision sciences points to this. Let's be clear that not all evidence contains data and not all data constitutes evidence concerning the issue at hand. Data may be relevant or it may not be. Hence the "wrong" or misappropriate date that is selected should be worthy of being maligned. Data is often not presented in ways that lets us judge if it is worthwhile or contains relevant data. Hence the dilemma – is it maligned or malignant? How much malaise or data fatigue is there in the work in which we engage? How do we plan accordingly? Strategically, for policy-makers, data is seen through preconceived notions, if it matches one's preconceptions, it is considered "good" data; otherwise it may be maligned.

Marcus Banks writes in 2010 how his interest in grey literature has shifted over the previous five years.[49] He speculates that grey data is compatible with grey literature due to the Web 2.0 directives and new opportunities to disseminate via many new avenues such as blogs, social media, tweeting. The end product or tweets has become "grey data" but we urge our audience and readers to consider the research by-products of both formal science and the social content or new methods to communicate it as worthy of retention and preservation. Libraries are

currently trying to influence governments and large enterprises to save, authenticate and preserve older runs or versions of data for future generations to compare, study and align with recent or current datasets, but that continues to be a challenge based on available resources, financial and human.

Malignant data can be construed as being misunderstood, clumsy or insufficiently analyzed – a sense of false positives, incomplete data, or a set of data that cannot be replicated using the same techniques and instruments for whatever reasons. This is increasingly problematic in the data sharing world. Funding sources may not want to consider supporting multiple efforts to study or engage in similar projects, however, it is the challenging outcome that may be providing clarity and a definitive analysis, not the original source. Building on previous work is the hallmark of research and taking it to the next level with more learned resourcefulness, an open and innovative mind makes it ultimately successful. Malignant data may still be vigorous but it is sick, perhaps suffering from malaise and fatigue and needs a new push to find its path.

Just two weeks ago, Pope Francis addressed the European Parliament with a grimly somber diagnosis about the state of Europe, that he described as a "continent's malaise," referring to how he perceives it to have lost its way, its energies sapped by economic crisis and a remote technocratic bureaucracy. It is increasingly a bystander in a world that has become "less and less Eurocentric," and that frequently looks at the Continent "with aloofness, mistrust and even at times, suspicion."[50] He went on to describe the aging product as "grandmotherly, no longer fertile and vibrant…and we cannot allow the Mediterranean to become a vast cemetery" and concluded another speech to the Council of Europe in Strasbourg on this same trip, by expressing more optimism, "It is my profound hope that the foundations will be laid for a new social and economic cooperation."[51]

We concur with O'Neil and Schutt that we would like to "encourage the "next-gen" data scientists to become problem solvers and question askers, to think deeply about appropriate design and process, and to use data responsibly and make the world better, not worse."[52] And to echo Pope Francis with his optimistic pledge for cooperation and not despair. That also makes data less maligned, malignant and grey.

References

[1] Auger, C.P., ed (1989). *Information Sources in Grey Literature, 2d ed.* London: Bowker-Saur.

[2] "Luxembourg Definition" of Grey Literature (http://opensigle.inist.fr/handle/10068/697932)

[3] Schopfel, J, Farace, D.J. (2010), "Grey Literature." In Bates, M.J., Maack, M.N. eds., *Encyclopedia of Library and Information Sciences, 3rd ed*.. Boca Raton, FL: CRC Press, 2029.

[4] OMB Circular 110 (http://www.whitehouse.gov/omb/circulars_a110#36)

[5] Data Science Initiative, University of California, Irvine (http://datascience.uci.edu)

[6] See http://datascience.uci.edu/about/

[7] Adapted from Defining Research Data (http://library.uoregon.edu/datamanagement/datdefinned.html) by the University of Oregon Libraries.

[8] Clarke-Stewart, K.A. (1989). "Infant Day Care: Maligned or Malignant?" *American Psychologist* 44, #2, 266-273.

[9] *Ibid*, 266.

[10] Data Paper (http://www.gbif.org/publishingdata/datapapers)

[11] Dominus, S. (2014). "You Can Affect Billions of People," *New York Times Magazine*, November 2: 14.

[12] Fienberg, Sharing Research Data (1985)

[13] Committee on Intellectual Property Rights and the Emerging Information Infrastructure, Computer Science and Telecommunication's Board Commission on Physical Sciences, Mathematics and Applications, and the National Research Council *(2000). Digital Dilemma: Intellectual property in the information age.* Washington, DC: National Academies Press. http://uclibs.org/PID/31331

[14] Internet Manifesto 2014. See http://www.ifla.org/publications/node/224

[15] eMail announcement from Martyn Wade, November 25, 2014.

[16] Elliott, S. (2014). "At Ogilvy, New Unit Will Mine Data," *New York Times*, October 21, 2014: B6.

[17] Lanser, J. (2014). "A Whole Lot of Data," *Orange County Register,* October 12. Real Estate Section: 6. 14.

[18] Westerman, G., Bonnet, D., and McAffee, A., (2014). *Leading Digital: Turning Technology into Business Transformation*. Boston, MA: Harvard Business Review Press, 113.

[19] See current job descriptions for Data Management positions such as at the University of Michigan (http://umjobs.org/job_detail/103342/research_data_curation_librarian or at the Pennsylvania State University Libraries (http://www.libraries.psu.edu/psul/jobs/facjobs/sdl.html) or at McGill University (http://www.mcgill.ca/library/files/library/14-al9904-01-coordinator-dat-curation-scholarly-communications.pdf) or at the University of Texas at Arlington (https://uta.service-now.com/jobs/recruiting_job_posting_detail.do?sysparm_document_key=u_recruiting_job_posting,a8f9b7706f6c75009f15cccc5d3ee4c3)

[20] Lohr,S. (2014). "For Big-Data Scientists, "Janitor Work' is Key Hurdle to Insights," *New York Times*, August 18: B4.

[21] See http://digitalcommons.iwu.edu/

[22] Davis-Kahl, S and Helmsley, M., eds. (2013). Common Ground at the Nexus of Information Literacy and Scholarly Communication. Chicago, IL: ACRL. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/commonground_oa.pdf

23 Coalition for Networked Information, Spring 2014 meeting, St. Louis, Mo. (http://www.cni.org/events/membership-meetings/past-meetings/spring-2014/)

[24] Data Science Initiative, University of California, Irvine (2014). (http://datascience.uci.edu/). Forum held October 24.

[25] Suber, P. (2008). Open-Access Timeline (http://legacy.earlham.edu/~peters/fos/timeline.htm

[26] O'Neil, C. and Schutt, R., (2013). Doing Data Science. Sebastopol, CA: O'Reilly Media.

[27] White House (2014) Big Data: Seizing Opportunities, Preserving Value (http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf)

[28] *Ibid.* Fact Sheet, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf

[29] Hurwitz, J., Nugent, A., Halper, F., Kaufman, M. (2013). *Big Data for Dummies*. Hoboken, NJ: Wiley: 280.

[30] *Ibid*. Fact Sheet.

[31] *Ibid*. Fact Sheet.

[32] Uprichard, E. (2014), "Big Data Doubts," *Chronicle of Higher Education*, October 17: B14-15.

[33] http://www.clearstorydata.com/

[34] http://www.statista.com/

[35] http://www.trifacta.com/platform/

[36] http://www.paxata.com/

[37] http://hortonworks.com/hadoop/mapreduce/

[38] http://www.dclab.com/

[39] Gross, M. (2014). "Data Conversion in the 21st Century," *Information Today*, October: 26.

[40] Taleb, N. (2013). "Beware the Big Errors of 'Big Data.'" *Wired*, February 8. (http://www.wired.com/2013/02/big-data-means-big-errors-people/)

[41] Huffington, A., (2014). *Thrive: The Third Metric to Redefining Success and Creating a Happier Life*. New York: Random House: 140.

[42] Pearson, K (1897). "Mathematical Contributions to the Theory of Evolution – On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs," Proceedings of the Royal Society of London, 60:489-498. DOI: 10.1098/rspl.1896.0076

Proceedings of the Royal Society of London **60**: 489–498. doi: 10.1098/rspl.1896.0076.

[43] Pawlowsky-Glahn, V. and Buccianti, A., eds. (2011). Compositional Data Analysis: Theory and Applications. Hoboken, NJ: Wiley.

[44] Spurious Correlations. http://www.tylervigen.com/

[45] https://www.lib.purdue.edu/researchdata

[46] http://cdluc3.github.io/dash/

[47] Ensuring the Integrity, Accessibility and Stewardship of Research Data in the Digital Age (2009). Washington, DC: National Academies Press.(http://www.nap.edu/penbook.php?record_id=12615)

[48] Bell, S. (2014). "MLD: Masters in Library Design, Not Science: From the Bell Tower." Library Journal, November 20, 2014. http://lj.libraryjournal.com/2014/11/opinion/steven-bell/mld-masters-in-library-design-not-science-from-the-bell-tower/

[49] Banks, M. (2010). "Blog Posts and Tweets: The Next Frontier for Grey Literature." In Farace, D., and Schopfel, J., eds., *Grey Literature in Library and Information Studies*. Berlin: DeGruyter Saur: 222-223

[50] Higgins, A. (2014). "At European Parliament, Pope Bluntly Critiques a Continent's Malaise," *NY Times,* November 26: A4, A10.

[51] *Ibid*.

[52] O'Neil and Schutt, chapter. 16.

Additional Resources:

Bessis, N. and Dobre, C., eds., (2014) *Big Data and Internet of Things: A Roadmap for Smart Environments*. Studies in Computational Intelligence, v.546. Cham: Springer.

Big Data Now: Current Perspectives from O'Reilly Media (2011). Sebastopol, CA: O'Reilly.

Chen, M., Mao, S., Zhang, Y., Leung, V, (2014). Big Data: Related Technologies, Challenges and Future Prospects. Cham: Springer.

Gauntner Witte, G. (2014). "Content Generation and Social Network Interaction within Academic Library Facebook Pages*," Journal of Electronic Resources Librarianship* 26 (2): 89-100.

Krier, L. and Strasser, CA (2014). *Data Management for Libraries: A LITA Guide*. Chicago, IL ALA TechSource.

Manning, P. (2013). *Big Data in History*. New York: Palgrave Macmillan.

Marcum, D, and George, G. (2010). *The Data Deluge: Can Libraries Cope with eScience*? Littleton, CO: Libraries Unlimited.

Pryor, G., Jones, S., and Whyte, A., eds. (2014). *Delivering Research Data Management Services: Fundamentals of Good Practice*. London: Facet.

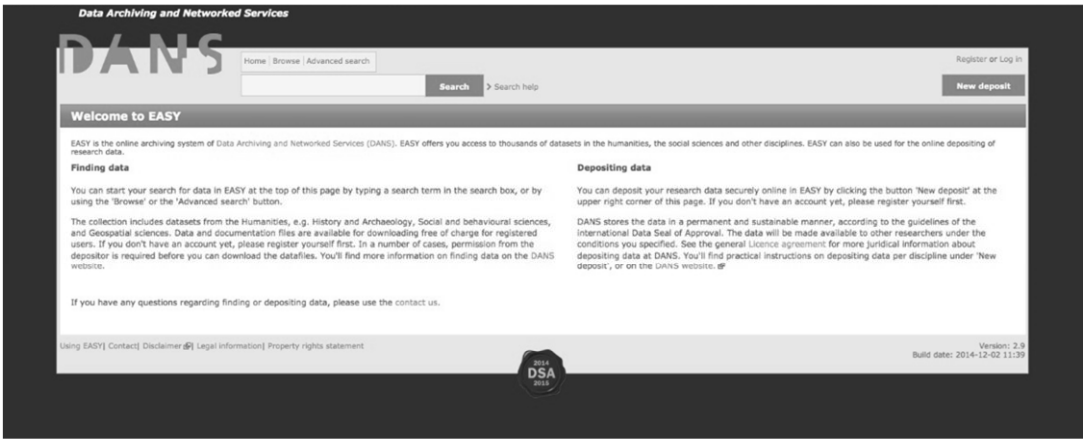
Open data, grey literature and disciplinary differences – Perspectives from a Dutch data archive

Marnix van Berchum
DANS, Data Archiving and Networked Services, The Netherlands

Abstract:
Since 2005 Data Archiving and Networked Services (DANS) promotes sustained access to digital research data. DANS offers several services to support this, including the online archiving system EASY, the Dutch Dataverse Network and the portal NARCIS. In this paper these services will be presented, including the differences we encounter between the disciplines using these services. Within the disciplines served by DANS – mostly belonging to the Social Sciences and Humanities – differences can be discerned regarding the 'openness' of the data. As case study the archeology datasets in EASY will be discussed. Many of them contain reports of archeological excavations done in the Netherlands. Should we consider these as research data or grey literature? And, should we open up the access to these datasets? And if not, why not? These, and other questions, will be addressed, providing a view on how DANS currently deals with open data, grey literature and disciplinary differences.

About DANS
DANS, Data Archiving and Networked Services[1], is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW)[2] and the Netherlands Organisation for Scientific Research (NWO)[3]. DANS is founded in 2005, as a 'merger' of existing institutions dealing with research data in the Netherlands. Taking into account these predecessors – the Steinmetz Archive (founded 1964) and the Historical Data Archive (founded 1989) – DANS builds on fifty years of experience in the storage and management of research data. DANS has as its mission to promote sustained access to digital research data. 'Digital research data' is understood to mean: research information, research data (such as databases, spreadsheets, text, images, audio, video, and multimedia) and digital publications (including preprints, reports). Although DANS is not primarily involved with Grey Literature, its services do hold a substantial amount of grey material, and as this paper will show, DANS supports the ideas of this community regarding the optimal availability of research data.

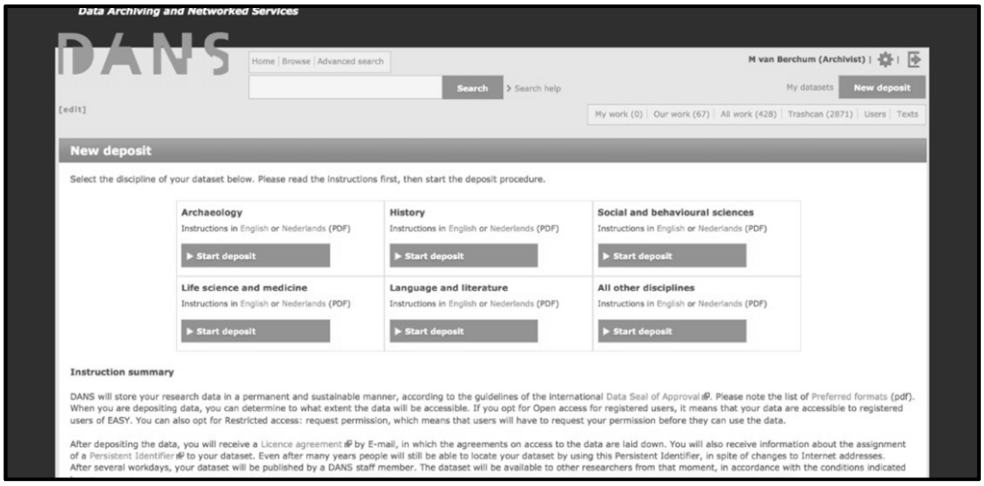
The Services of DANS
To accomplish its mission of the promotion of sustained access to research data, DANS offers three applications for the management, storage and dissemination of research data, and several non-technical services (e.g. training and consultancy, and support on the certification of digital repositories). Below a short introduction on the data services is given.



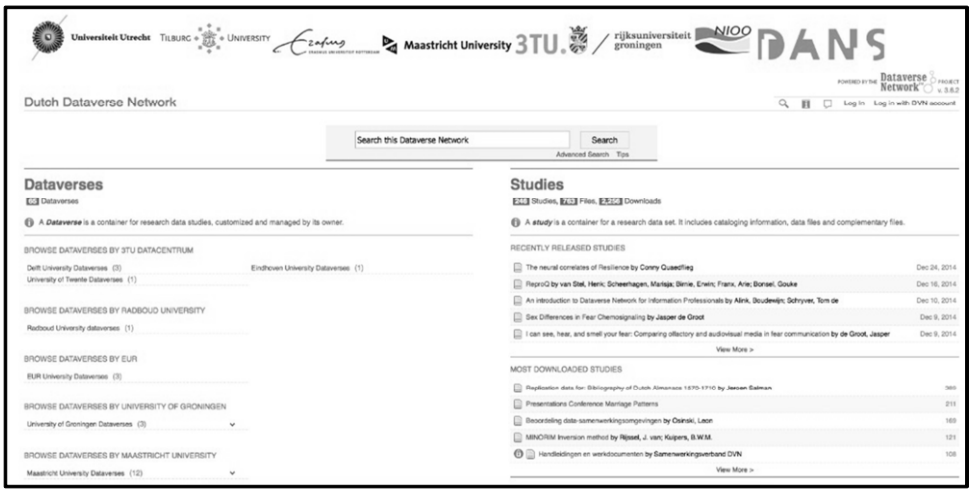
Picture 1. – Homepage of the EASY service, www.easy.dans.knaw.nl.

EASY: Certified Long Term Preservation of Research Data
EASY (Electronic Archiving SYstem) is the system of DANS for the long-term preservation of research data. At the time of writing it hosts a rough 35 thousand datasets, mainly related to the disciplines Humanities and Social Sciences. EASY is a Data Seal of Approval (DSA) certified

repository[4], suitable for the storage of data as required by e.g. the Dutch governmental funder NWO[5]. Individual users are able to manually upload their data, including the appropriate metadata. Furthermore, EASY has several more automated processes of ingesting content into the system, connected to other systems (e.g. local data repositories or Virtual Research Environments, hosted at DANS' client institutions).
EASY offers six deposit procedures, tailored to specific disciplinary requirements. These may include additional metadata, like mandatory metadata fields for the Dutch Archeology community[6] (see also below) or the CMDI metadata format of the CLARIN research infrastructure for language and linguistics data[7].



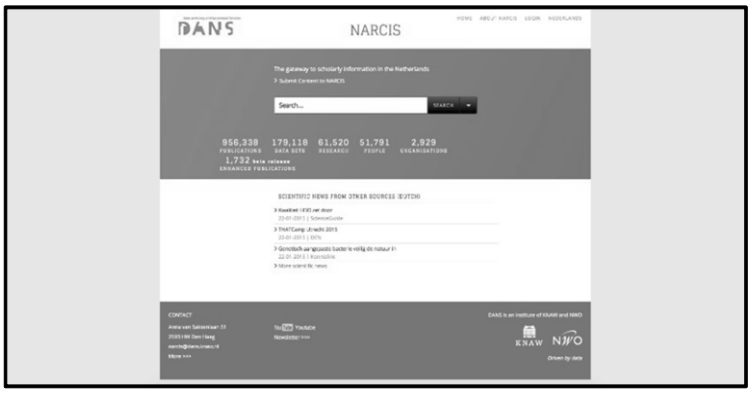
Picture 2. – The six disciplinary deposit entries of EASY.



Picture 3. – Homepage of Dutch Dataverse Network (DDN), www.dataverse.nl.

Dutch Dataverse Network: storing data during the research processes

The Dutch Dataverse Network (DDN) is a collaboration of DANS and nine Dutch institutions of Higher Education[8]. DDN started as a project at Utrecht University, in which the university library wanted to explore the possibilities of hosting and managing research data at the institutional level. Several other university libraries joined this initiative, and in a close cooperation between the two institutions, Dutch Dataverse Network was migrated from Utrecht University to DANS in Spring 2014.
In recent years the Research Data Netherlands (RDNL) consortium[9] introduced the so-called *Front Office | Back Office model*, in which the different functions and roles in the data management process are divided. The RDNL partners supply the Back Office functionalities, like long-term preservation of data and training and consultancy for the Front Offices. The latter, e.g. the university libraries, support the researchers and host data application(s) for the duration of the research and mid term storage. DDN is well suited to take this role, offering the Front Offices a tool with which they can control their institutional research data management requirements.



Picture 4. – Homepage of NARCIS, www.narcis.nl.

NARCIS: the gateway to scholarly information in the Netherlands

While EASY and DDN deal with research *data*, the NARCIS portal has a broader scope. NARCIS provides access to different types of scholarly output, including (open access) publications from the repositories of all the Dutch universities, KNAW, NWO and a number of research institutes, and descriptions of research projects, researchers and research institutes. All datasets from EASY can also be found via NARCIS, as well as datasets from other Dutch data archives (like 3TU.datacentrum and The Language Archive). NARCIS harvest the metadata of the abovementioned resources from the institutions, and is at the same time the hub for harvesting the aggregated metadata by service like Google Scholar and OpenAIRE[10].

Open Access at DANS

DANS – as well as its host institution KNAW and NWO – supports the Open Access principles, and stimulates researchers to make their scholarly output available as wide as possible. Acknowledging that data cannot always be open, e.g. in the case of sensitive data containing personal information, the motto of DANS is: "*Open if possible, restricted if necessary!"*

All metadata of the data sets in EASY are Open Access available to all users, and are also harvestable via an OAI-PMH endpoint. In 2014 DANS changed the access categories in EASY, for the first time offering a full Open Access license to the data files. The current categories, with which a dataset can be deposited, are[11]:

1. Open access
2. Open access for registered users
3. Restricted access
4. Other access

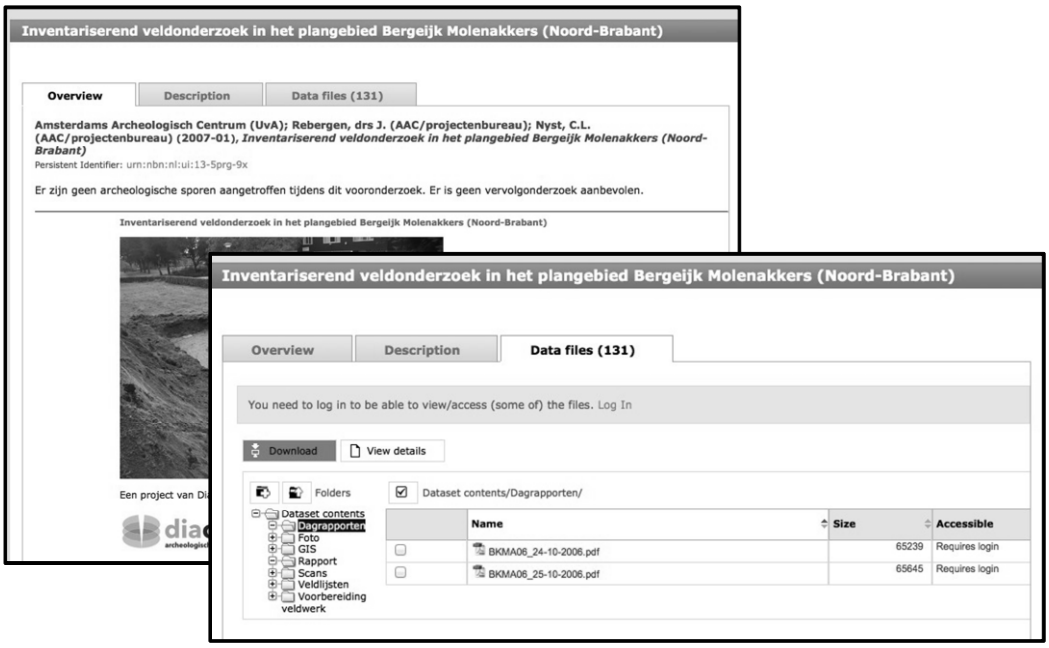
The first option is a *CC0 No Rights Reserved* licence[12], making it possible for human and machine users to freely use and re-use the data. Although the CC0 licence does not require an attribution to the creator of the data when re-using, like the related CC BY licences do, this should not keep the scholarly user from properly citing the data she uses[13]. To this end DANS supports data citation with the DataCite Digital Object Identifier (DOI)[14].

Since the implementation of the new CC0 option, just a few datasets have this access category. DANS is actively promoting depositors to deposit with this new category, or reassign existing datasets to this category.

Grey at DANS

DANS is in several ways connected to grey literature. First of all, DANS hosts the datasets related to the Grey Literature Conference Proceedings[15]. All authors of papers at the Greynet Literature Conference are allowed, free of cost, to deposit their related data or supplementary material in EASY. Furthermore, the data documentation files and publications deposited together with the data files can be considered to be grey literature as well.

NARCIS, as national aggregator of the Dutch higher education institutions, contains a substantial amount of metadata referring to grey literature. An estimated 33% of the contents of NARCIS can be classified as grey, existing of reports, unpublished doctoral theses, conference papers etc. present in the institutional repositories.



Picture 5. – Views of an archeological dataset in EASY.

The archeological datasets

A special type of dataset in the EASY archive has a shade of grey as well: the datasets deposited by the archeological community. All archeologists in the Netherlands are legally obliged to digitally deposit the results of their excavations results in EASY. Special ingest procedures are devised, in which an archeologist can deposit her material at DANS and, at the same time, register the excavation at the governmental institution involved[16]. Due to this obligation, archeology is a well-represented discipline in EASY, occupying a rough 80% of all data.

A typical archeological dataset consist of the excavation report, pictures taken and geographical data on the location. Disciplinary additional metadata fields are needed and include e.g. references to the excavation number[17]. To preserve the historical buildings and archaeological sites, and perform the archeological research as effectively as possible, most archeological datasets are available under the 'Restricted Access – Archeology Group'. This means only archeologists registered at DANS are able to access the files present in the dataset. As mentioned earlier, all metadata is freely available to all, making these grey resources available to a global audience.

References

DANS Licence Agreement, 2014, http://dans.knaw.nl/en/about/organisation-and-policy/legal-information/DANSlicenceagreementUK.pdf

Dillo, I., and Leeuw, L. de, *Data Seal of Approval: Certification for sustainable and trusted data repositories*, 2014 Data Archiving and Networked Services (DANS), http://www.datasealofapproval.org/media/filer_public/2014/10/03/20141003_dsa_overview_defweb.pdf.

A federated data infrastructure for the Netherlands: the front-office - back-office model, 2014 Research Data Netherlands (RDNL), http://www.researchdata.nl/fileadmin/content/RDNL_algemeen/Documenten/RDNL_FOBOmodel-UK-web.pdf.

Hollander, H., 'The e-Depot for Dutch Archaeology: Archiving and publication of archaeological data', in *Proceedings of the 18th International Conference on Cultural Heritage and New Technologies 2013*, 2014 Museen der Stadt Wien – Stadtarchäologie, http://www.chnt.at/wp-content/uploads/Hollander_2014.pdf.

Riding the wave: How Europe can gain from the rising tide of scientific data, 2010 High Level Expert Group on Scientific Data, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf.

Sustained access to digital research data: Summary of the DANS Strategy Policy 2011-2015, 2011 Data Archiving and Networked Services (DANS), http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material/DANS-strategy-policy-20112015-summaryUK.pdf.

Wiersma, P., *Podium Plus: Onderzoek naar de mogelijkheden tot landelijke samenwerking bij de opslag van onderzoeksdata [Podium Plus: Exploration into the possibilities of a national, cooperative storage of research data],* 2011 SURF, https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2011/Podium+Plus+-+definitief.pdf.

[1] www.dans.knaw.nl.
[2] www.knaw.nl.
[3] www.nwo.nl.
[4] For more information on the Data Seal of Approval see www.datasealofapproval.org; the assessment document of the DANS-EASY seal can be found at https://assessment.datasealofapproval.org/assessment_101/seal/pdf/.
[5] See http://www.dans.knaw.nl/en/deposit/information-about-depositing-data/data-contract/data-contract?set_language=en and http://www.nwo.nl/en/news-and-events/dossiers/datamanagement.
[6] https://easy.dans.knaw.nl/doc/DANSarchaeologicalmetadataUK.pdf.
[7] http://www.clarin.eu/content/component-metadata.
[8] At the time of writing these are the three technical universities of the 3TU consortium (Eindhoven University of Technology, University of Twente and TU Delft), Erasmus University Rotterdam, Maastricht University, NIOO – Netherlands Institute of Ecology, Tilburg University, University of Groningen, and Utrecht University.
[9] www.researchdata.nl. RDNL is a collaboration of DANS, 3TU.datacentrum and SURFsara.
[10] See http://www.narcis.nl/background/Language/en for an overview of the content of NARCIS. For a full list of organizations or services that harvest NARCIS, see http://www.narcis.nl/dataprovider/Language/en.
[11] See the Appendix, for a full description of the categories available.
[12] https://creativecommons.org/about/cc0.
[13] Cf. *The Netherlands Code of Conduct for Academic Practice*, as published by the Association of universities in the Netherlands (VSNU), at www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The%20Netherlands%20Code%20of%20Conduct%20for%20Academic%20Practice%202004%20%28version%202014%29.pdf.
[14] https://www.datacite.org. The DOI identifier will be implemented in DANS-EASY in the first half of 2015.
[15] http://www.greynet.org/dansdataarchive.html.
[16] This collaboration of DANS with the Cultural Heritage Agency (RCE) is formally known as the e-Depot for Dutch Archaeology (EDNA), www.edna.nl.
[17] See https://easy.dans.knaw.nl/doc/DEP_ArchaeologyUK.pdf and https://easy.dans.knaw.nl/doc/DANSarchaeologicalmetadataUK.pdf.

Appendix – description of the Access rights available in EASY

[The text below is taken from the help text in EASY (visited 22 January 2015)]

1. Open access
The files are accessible to all users of EASY.

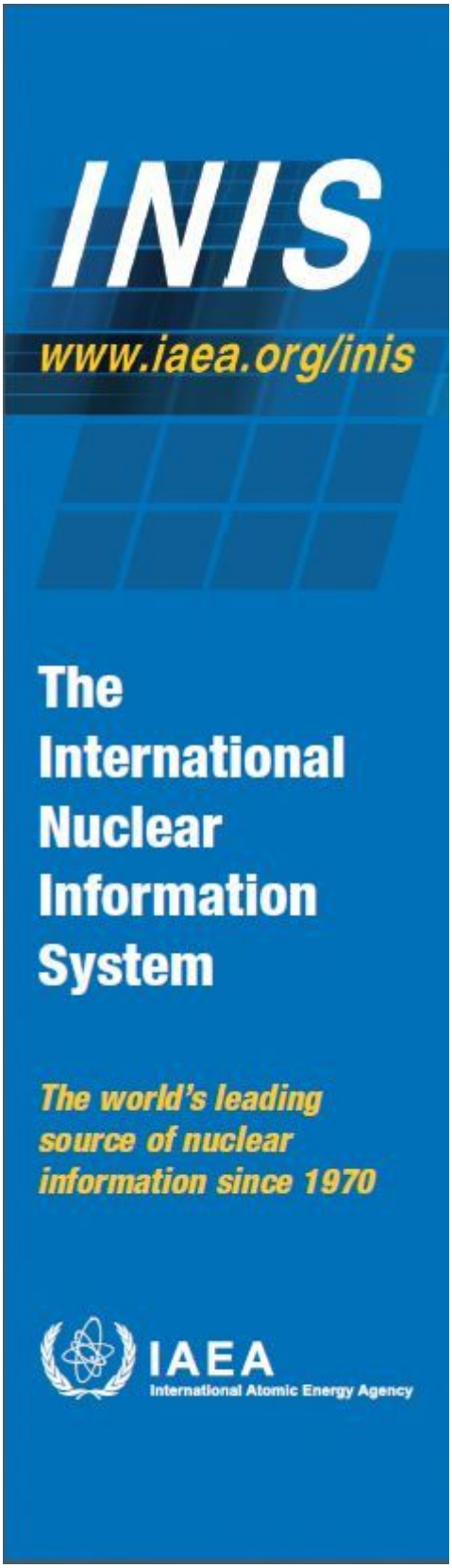
2. Open access for registered users
The files are accessible to all registered users of EASY.

3. Restricted access
The access to the files deposited is limited to a smaller group of users. EASY currently has the two following forms of limited access:

- Restricted - archaeology group.
Please note that this option can only be selected for archaeological deposits.
The datasets are only accessible to registered users of EASY who belong to a specific group. In principle, this option is used as default for all archaeological datasets. Just like Archis is an information system for and by archaeologists, the archaeological files in EASY are also only accessible to other, registered archaeologists. We share the available information on archaeological sites, without this information being immediately available on the Internet so that we can preserve the historical buildings and/or archaeological sites or perform scientific research as effectively as possible.

- Restricted - request permission.
With this option, you - as the person depositing the dataset - can personally manage the accessibility to the files deposited by you. The datasets are accessible only if another registered user, not necessarily an archaeologist, has submitted a 'permission request'. You, the person depositing the dataset, will be informed of this request and you will personally be able to grant or reject this request for permission to access the dataset.

4. Other access
It is not possible to access the datasets through EASY. DANS is only intended for sustainable, digital archiving behind the scenes. Another possibility is that the data have been stored elsewhere and that only the metadata have been stored in EASY. You can only choose this option if the digital data are accessible, for example, through a different data repository, digital archive, databank or website.


INIS
www.iaea.org/inis
The International Nuclear Information System
The world's leading source of nuclear information since 1970
IAEA
International Atomic Energy Agency

Enhancement of the functions of the Japan Atomic Energy Agency Library's Fukushima Nuclear Accident Archive using a novel data flagging system that improves the utilization of numerical data on the Internet

Kiyoshi Ikeda, Mayuki Gonda, Shun Nagaya, Misa Hayakawa, Yukinobu Mineo, Katsuhiko Kuni, Minoru Yonezawa, Keizo Itabashi
JAEA, Japan Atomic Energy Agency, Japan

Abstract
Related to the Fukushima Daiichi Nuclear Power Station Accident, the Japan Atomic Energy Agency Library has accumulated valuable information on Internet. In the Fifteenth International Conference on Grey Literature, we reported the development of the Fukushima Nuclear Accident Archive using the DSpace. We have encountered a new, challenging issue of grey literature. In many cases, Internet information contains valuable numerical data. However, identifying the existence of numerical data on Internet sites is difficult, and the metadata created for the Fukushima Nuclear Accident Archive cannot currently be used to distinguish whether such information contains numeric data. Therefore, we have considered a method to identify numerical data and have introduced a "data flagging" system that has been used in the International Atomic Energy Agency's International Nuclear Information System. In this paper, we introduce the proposed data flagging system and discuss its application to the Fukushima Nuclear Accident Archive.

1. Introduction

The Fukushima Daiichi Nuclear Power Station Accident (hereafter referred to as "the Fukushima Accident") occurred in March 2011. Information about the accident has primarily been published via the Internet, and permanent accessibility to the information, bibliographic control, and so on have been critical issues in the collection, preservation, and provision of Internet information related to the Fukushima Accident. We previously reported the development of an archive using DSpace to solve such issues and have discussed "Internet information is Grey Literature" as a new type of grey literature at the Fifteenth International Conference on Grey Literature [1].

A widely diverse and massive amount of data is utilized on the Internet. In many cases, Internet information contains valuable numerical data, such as observed, compiled, evaluated, experimental, statistical, and theoretical data.

However, determining the existence of numerical data on Internet sites and distinguishing whether information contains numeric data using our created metadata for the archive is difficult. An effective method to identify whether information contains such data would be very convenient.

We have considered a method to identify numerical data and attempted to introduce a "data flagging" system, which has been used in the International Atomic Energy Agency's International Nuclear Information System (INIS). In the INIS, data flagging is used to identify the existence of useful data in documents and literature.

In this paper, we introduce our data flagging system for numerical data on the Internet and its application to the Fukushima Nuclear Accident Archive.

2. Status of the Fukushima Nuclear Accident Archive

2.1 Collection of Fukushima Accident Information

The Japan Atomic Energy Agency (JAEA) Library is one of the largest nuclear information centers in Japan. After the Fukushima Accident, we have supported scientific research and development activities related to the Fukushima Accident, and the archive is one of the major missions of the JAEA Library [2][3]. Since April 2011, we have collected, compiled, and distributed Fukushima Accident's reference information using a special website [Fig. 1].

This website consists of

(1) research and development results related to the Fukushima Accident by JAEA staff,

(2) bibliographies related to the Three Mile Island Unit 2 reactor, Chernobylsk-4 reactor accident, environmental decontamination, and other topics (19 themes, approximately 18,000 articles), and

(3) Internet sources related to the Fukushima Accident (30 fields, approximately 4,000 URLs).

Fig. 1 JAEA Library website related to the Fukushima Accident

Another role of the JAEA Library is to act as an archive for the INIS National Center of Japan. The INIS is an open access database of published scientific literature on the peaceful uses of nuclear science and technology. The JAEA Library has provided database input records for literature (books, articles, conference proceedings, and technical reports) related to the Fukushima Accident to the INIS.

However, much of the information related to the Fukushima Accident, such as records, documents (press releases and reports), photos, videos, and numerical data (e.g., radiation monitoring and plant data), is published via the Internet, which makes an entire coverage difficult. Parts of this Internet information has been distributed as grey literature because of a lack of bibliographical control and some websites not having permanent accessibility.

Thus, in 2013, we began to collect, organize, and disseminate such information because preserving the observations and knowledge obtained from the Fukushima Accident is very important.

2.2 Outline of Fukushima Nuclear Accident Archive

The Fukushima Nuclear Accident Archive is a bibliographic database that covers Internet information of Japanese government-based organizations and the Tokyo Electric Power Company (TEPCO), as well as bibliographic information about oral presentations regarding credible nuclear associations or societies related to the Fukushima Accident [Fig. 2]. The Fukushima Nuclear Accident Archive was opened to the public on June 23, 2014, to support related research and development [Fig. 3] [4][5].

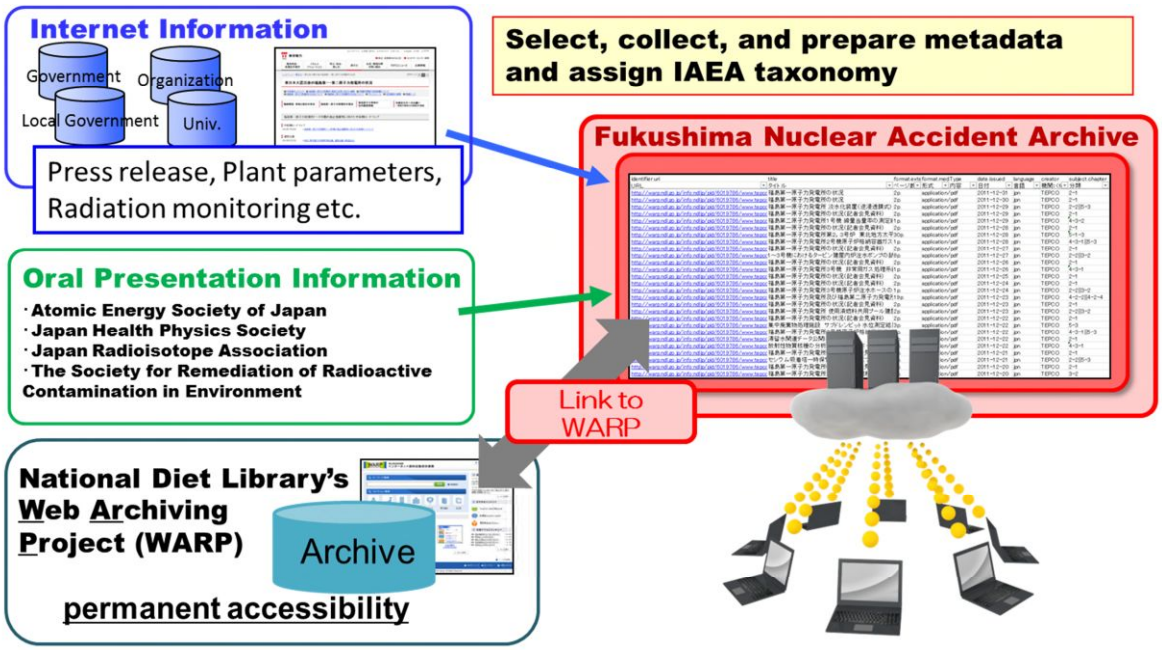


Fig. 2 Outline of the Fukushima Nuclear Accident Archive

To ensure permanent access to this Internet information, we adopted the National Diet Library's Web Archiving Project (WARP) [6]. The WARP stores official Japanese institutions' websites, such as those related to the federal government, the Diet, the courts, local governments, independent administrative organizations, and TEPCO.

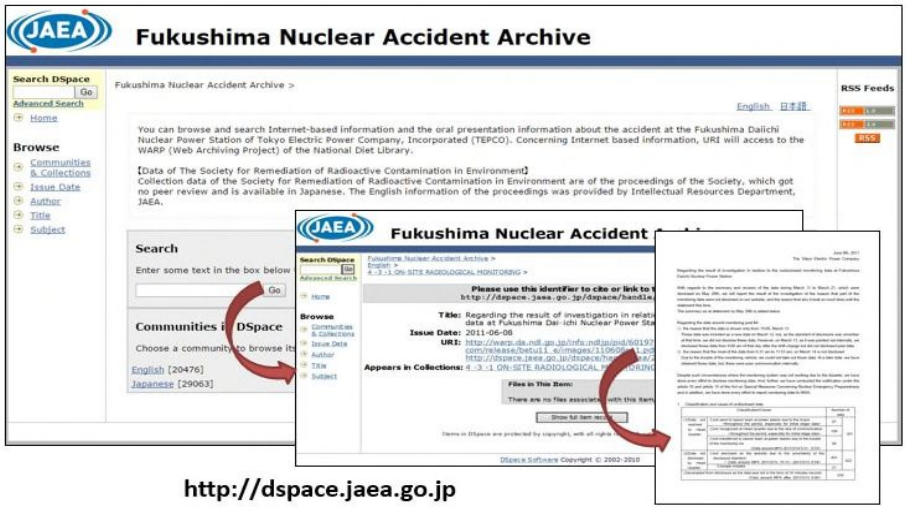


Fig. 3 Fukushima Nuclear Accident Archive

In the archive, we employ DSpace as the archive system because it is commonly used as an institutional repository system. DSpace supports the Open Archives Initiative Protocol for Metadata Harvesting (hereafter referred to as "OAI-PMH") and is capable of processing standardized metadata.

In addition, we also employ the Dublin Core Metadata Elemental Set (hereafter referred to as "Simple Dublin Core") and use the IAEA's Nuclear Accident Knowledge Taxonomy as a classification system.

We provide more than 50,000 metadata for Internet information related to the Fukushima Accident, such as records, documents, photos, videos, and numerical data, from agencies such as TEPCO; the Ministry of Economy, Trade and Industry; the Ministry of Education, Culture, Sports, Science and Technology; and the Ministry of the Environment.

3. Data Flagging System to Utilize Internet Information with Numerical Data

3.1 Numerical Data

Tremendous amounts of numerical data exist that have been used in various fields of scientific research and society. Numerical data refer to physical quantities, such as densities, melting points, cross sections, and spectral shifts, or the number of occurrences of certain events, such as failures of a piece of equipment and coefficients or parameters (such as the rate constant of a particular chemical reaction).

In the nuclear field, Internet information contains valuable and useful data, e.g., radiation monitoring data, plant parameters, physical quantities, and processed data, such as compiled, evaluated, experimental, statistical, and theoretical data.

Numerical data on the Internet has three issues.

(4) Identifying its existence is difficult.
(5) The structural aspects of the data, such as file format, content type, and whether it is raw or processed data, must be determined.
(6) The validity and reliability of numerical data acquisition methods require evaluation.

In many cases, no information exists in the metadata that indicates whether the data contains numerical data. In addition, information often exists in a mixed state of various types of documents (photos, videos, text, and so on) [Fig. 4]. Thus, the metadata must identify whether such data contains numerical data.

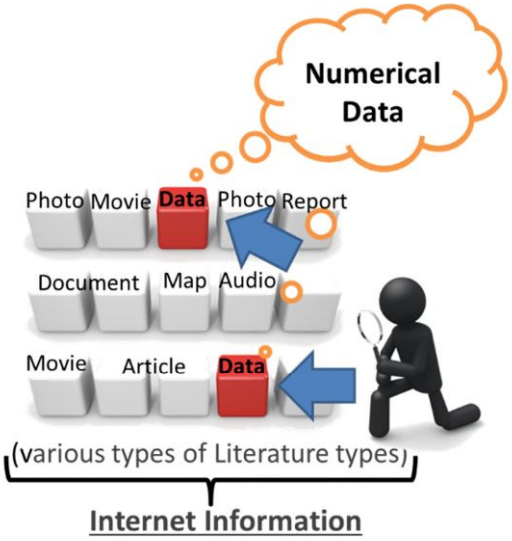


Fig. 4 Various types of information

Another issue is related to the reuse of information. For example, numerical data in PDF or JPEG format is inconvenient; it must be converted to comma-separated value (CSV) format by processing for effective use. In addition, the quality of the information is often an issue. The latter two issues can only be addressed after the information has been obtained. In this paper, we introduce efforts to solve the first issue.

For many people engaging in research and development, availability of a useful method to identify whether information contains useful data would be very advantageous. Thus, we introduce a data flagging system to identify numerical data in a document.

3.2 Data Flagging

Data flagging identifies the existence of useful numerical data in documents and literature. Data flagging has been used in the INIS as a mechanism to create bibliographical data [7].

Generally, documents should be data flagged by performing two steps [8]: 1) add one or more of the narrower terms of the descriptor NUMERICAL DATA to the already chosen set of descriptors

and 2) place the Literary Indicator tag "N," which emphasizes certain literary characteristics of a record, in the appropriate place within the bibliographic data elements.

We have considered such a method to identify numerical data and have attempted to implement a data flagging system in the INIS because data flagging is effective in searching valid numerical data. However, currently, using keywords in the INIS database is not possible. To implement data flagging to the Fukushima Nuclear Accident Archive, we must consider appropriate keyword methods and use tags that indicate numerical data.

3.3 Keywords for Data Flagging

Keywords for data flagging are very useful in the identification of numerical data. We have considered using keywords in the INIS database and the INSEPC database [9]. These databases each have a control thesaurus of terms related to nuclear and physics fields and numerical data indexing rules that include data flagging: categorized keywords (e.g., compiled, evaluated, experimental, and theoretical data.) in the INIS database, and physical quantity keywords (e.g., temperature, pressure, radiation exposure, and radioactivity) in the INSPEC database.

However, there is a lack of keywords closely related to the Fukushima Accident. If we only make use of both thesauruses' keyword, we must add additional terms related to the Fukushima Accident. To address this lack of terms, we used the content of *Plant Data of Fukushima Daiichi Nuclear Power Station at the time of the Tohoku-Chihou-Taiheiyou-Oki Earthquake* [9] because this publication contains a guide to relevant data about the Fukushima Accident, e.g., plant parameters, data of abnormal events (including alarm records), and transient recorder data.

We have also checked the data in the Fukushima Nuclear Accident Archive because these data are closely related to the Fukushima Accident, such as water level, status of water leakage, radioactivity level, reactor vessel pressure, monitoring, and water injection into the reactor.

We considered the keywords listed in Fig. 5 for data flagging by combining the thesaurus **keywords** of the INIS and INSEPC databases and the content and terms extracted from the Fukushima Nuclear Accident Archive.

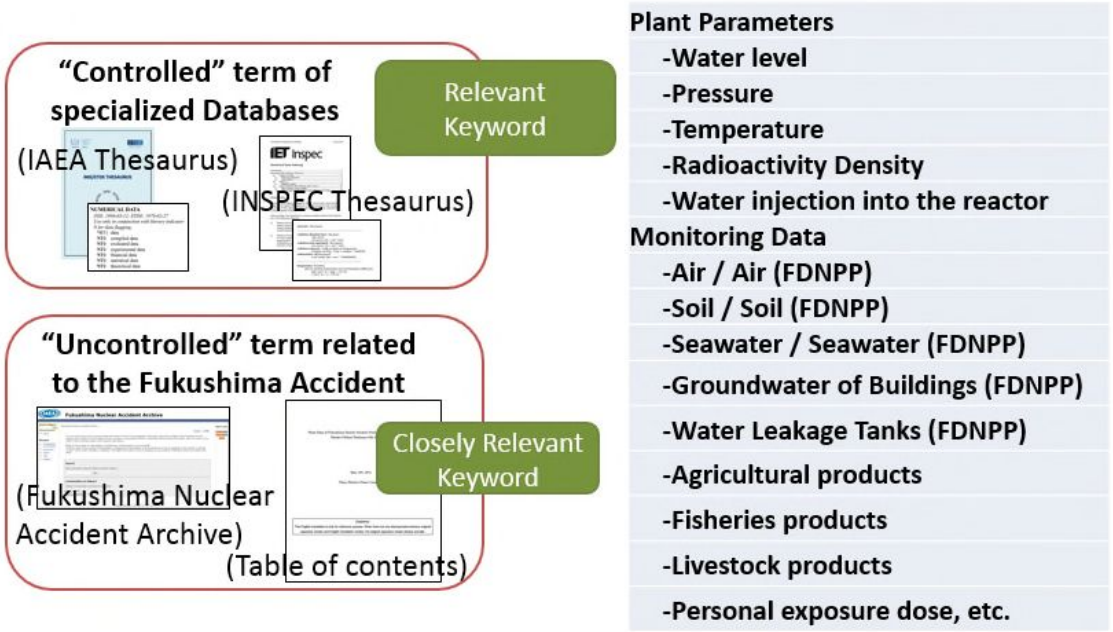


Fig. 5 Consideration of the keywords for data flagging on the basis of the two databases' thesaurus

3.4 Data Flagging in the Fukushima Nuclear Accident Archive

We use Simple Dublin Core as our metadata format and employ the bibliographical control rule for Internet information in the Fukushima Nuclear Accident Archive.

We use the following for data flagging [Fig. 6]:

- We assign "Numerical data" to the Type field, which indicates that the record contains numerical data.
- We assign keywords to the Subject.chapter field that indicate the type of numerical data.

Element	Value (ex)
Identifier.uri	http://warp.ndl.go.jp/info:ndljp/pid/3531775/www.nisa.meti.go.jp/english/files/en20110503-2-4.pdf
Title	Fukushima Dai-ichi Monitoring Data 2011-05-03 17:00
Format.extent	4p.
Format.medium	application/pdf
Type	Add Tag: "Numerical Data"
Date.issued	2011-05-03
Language	eng
Creator	METI
Subject.chapter	Add Keywords: "Monitoring Data-Air" "Monitoring Data-Air (FDNPP)" etc.

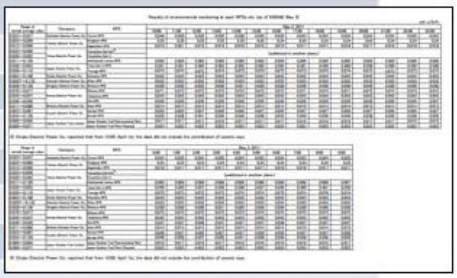


Fig. 6 Data flagging in the Fukushima Nuclear Accident Archive

Utilizing the tags and data flagging keywords, narrowing search results to those that contain numerical data is possible, which is beneficial for people who would like to access only numerical data.

3.5. Using Numerical Data Related to the Fukushima Accident

We face three issues in the development of effective metadata for numerical data related to the Fukushima Accident.

First, to create effective metadata, we must add multiple dates or times, the location of the observed or collected data, and distinguish between raw data and processed data. As a result, data can be obtained more accurately. However, despite using Simple Dublin Core to describe date information, multiple dates, such as date created, date modified, date collected, and date evaluated, cannot be distinguished because only one date field exists.

Second, to increase the convenience of the numerical data, conversion of various file formats (e.g., PDF, images, and graphs) to CSV or text format is required. However, this requires significant time and effort because such file formats represent non-structured data, the format of the data is frequently scattered, and the data is often not represented according to a unified format rule after initial conversion.

Third, we have no criteria to evaluate the validity and reliability of numerical data. In future, we plan to continue working on these challenging issues.

4. Conclusion

The amount of valuable numerical data related to the Fukushima Accident is increasing day by day, especially on the Internet. However, identifying whether such information contains numerical data even though we use search engines (e.g., Google) is difficult because numerical data can be provided in image files that are not indexed and websites do not always contain words that indicate numerical data.

We attempted to introduce a novel data flagging system to the Fukushima Nuclear Accident Archive to facilitate effective search for numerical data. Thus, users can significantly benefit by appropriate identification of the content of Internet information related to the Fukushima Accident.

We believe that our work will improve access to numerical data related to the Fukushima Accident and will hopefully contribute to improved circulation of grey literature from Japan to the world.

References

[1] Ikeda, K., Ohshima, T., Gonda, M., Nagaya, S., Hayakawa, M., Kunii, K., Mineo, Y., Yonezawa, M., and Itabashi, K. (2014) 'Contribution to the improvement of dissemination of grey literature -JAEA Library's efforts for collecting, organizing and disseminating information on nuclear accidents- 'Proceedings of the Fifteenth International Conference of Grey Literature, pp.41-47.

[2] Ikeda, K., Nakajima, H. and Yonezawa, M. (2012) 'Introduction of Useful Bibliographies and Resources related to the Fukushima Daiichi Nuclear Power Station Accident through the JAEA Library', ATOMOSU (ATOMO-sigma, Journal of the Atomic Energy Society of Japan), Vol.54, No.8, pp.549-553 (in Japanese).

[3] Nakajima, H., Ikeda, K., Yonezawa, M., Itabashi, K., Kiriyama, E. and Iwata, S. (2012) 'Study for development of the Fukushima Dai-ichi nuclear power station accident archive', Journal of Japan Society of Information and Knowledge, Vol.22, No.4, pp.344-353 (in Japanese), https://www.jstage.jst.go.jp/article/jsik/22/4/22_22_344/_pdf, retrieved 30 June 2013.

[4] Gonda, M., Ikeda, K., and Nagaya, S. (2014) 'Road to archiving of information relating to the Fukushima Nuclear Accident by the Japan Atomic Energy Agency Library(<Special feature>Earthquake disaster archives)', The journal of Information Science and Technology Association, Vol.64, No.9, pp.357-360 (in Japanese).

[5] Fukushima Nuclear Accident Archive. http://dspace.jaea.go.jp/dspace/?locale=en

[6] NDL Web Archiving Project (WARP).http://warp.da.ndl.go.jp/info/WARP_en.html

[7] INIS Collection Search. http:// http://inis.iaea.org/search/

[8] INIS/ETDE Manual for Subject Analysis (2012) IAEA-INIS/ETDE-03, International Atomic Energy Agency.

http://www.iaea.org/inis/products-services/publications/IAEA-INIS-ETDE-03/IAEA-INIS-ETDE-03-2012-08.pdf

[9] INSPEC. http://thomsonreuters.com/inspec/

[10] Plant Data of Fukushima Daiichi Nuclear Power Station at the time of the Tohoku-Chihou-Taiheiyou-Oki Earthquake, Tokyo Electric Power Company, May 16th, 2011.
http://warp.da.ndl.go.jp/info:ndljp/pid/6019786/www.tepco.co.jp/en/nu/fukushima-np/plant-data/f1_1_Setsumei.pdf

GreyGuide

Guide to Good Practices and Resources in Grey Literature

A Community Driven Open Resource Project in Grey Literature

"GreyGuide offers information professionals, practitioners, and students common ground for good practice in grey literature"

In December 2013, the GreyGuide Project was formerly launched as an online forum and repository of good practice in grey literature.

Open Source Repository of good practices in the field of grey literature. That which originated in monographic form is now open and expanded to include content from the global grey literature community

The GreyGuide will provide a unique resource in the field of grey literature that is long awaited and which responds to the information needs of a diverse, international grey literature community.

GreyNet's Web Access Portal

2013

Repository
of Proposed and Published
Good Practices
in Grey Literature

2014

New Collections
GL16 proposals
BIO notes
Grey Source Index

GG Portal
GreyNet's
Web Access Portal

PARTNERS

CNR ISTI
http://www.isti.cnr.it

GreyNet International
http://www.greynet.org/

PROJECT TEAM

Stefania Biagioni
Carlo Carlesi
Roberto Ponti & Chiara Bedini
ISTI-CNR
Italy

Joachim Schopfel
University of Lille
France

Christiane Stock
INIST-CNRS
France

Dominic Farace
GreyNet International
Netherlands

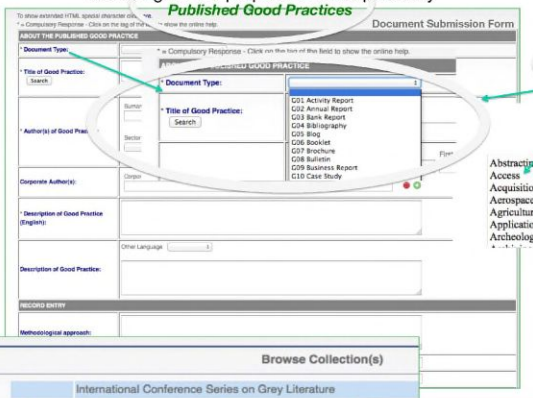
Welcome to the GreyGuide, your point of access to Grey Literature ***http://greyguide.isti.cnr.it***

Repository Key Features

- The system provides services to support the submission, description, searching, browsing, retrieval, access, preservation and visualization of documents
- Different search/browse options are offered: Google-like or fielded
- Full compliant Open Access implementation via the OAI-PMH protocol

Digital Library Technical Approach

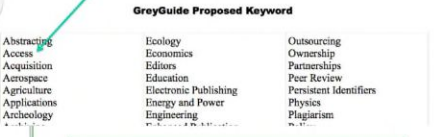
- The Repository is implemented on top of the Octopus Digital Library Management System
- Uses Dublin Core Qualified (DCQ) encoding for the purpose of interoperability



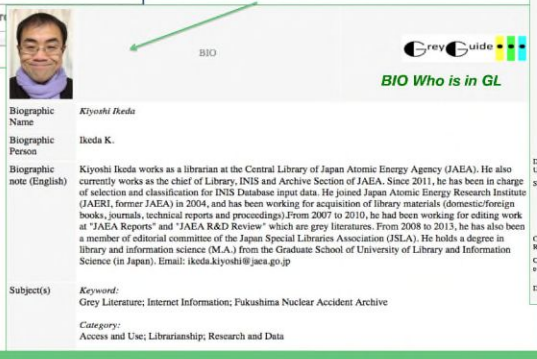
Repository User Interface

The Repository's User Interface has been designed to reflect the most recent advances in UI usability:

- The look and feel let users have an extremely comfortable access to functionality and content
- All most common functions can be easily accessed via one-click commands



....to enable contributors and users the opportunity to identify areas of good practice related to grey literature, to facilitate in assigning descriptive terms during record entry, and to provide users with relevant terms in their search for good practices in grey literature....



http://greyguide.isti.cnr.it

Remember to endorse The Pisa Declaration!

COLLECTIONS

CURRENT ...

2013
- *Published Good Practices*
- *Proposed Good Practices*

2014
- *Conference Proposals*
- *Who is in Grey Literature*
- *GreySource Index,*

... IN PROGRESS

2015
- *GreyForum Series*
- *International Directory of Organizations*
- *GreyNet Inhouse Publications*

CONTACT
Info@greynet.org

GreyGuide, GreyNet's web access portal and lobby for change in Grey Literature

Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands
Stefania Biagioni, **Carlo Carlesi and Roberto Ponti,** ISTI-CNR, National Research Council, Italy
Christiane Stock, Inist-CNRS, National Centre of Scientific Research, France

Introduction

In December 2013, the GreyGuide was formerly launched as an online forum and repository of good practice in grey literature. The project partners then turned to the acquisition of both proposed and published good practices. During this same timeframe, GreyNet – one of the project partners – welcomed far reaching developments in its infrastructure. Three new committees were established alongside its Program Committee in line with GreyNet's fourfold mission dedicated to research, publication, open access, and education in the field of grey literature.

In the process of coordinating and facilitating the work of these new committees, it became clear that a multitude of web-based content that is currently maintained on GreyNet's website and conference site - accessible on diverse webpages in PDF format - could better be made accessible via a web portal. This would allow for browse, search, and retrieval across resources and collections. The GreyGuide was tested for this purpose and it was then decided to select from GreyNet's range of content and commence with migration to the GreyGuide. While the web origins of three such collections were soon identified – namely, the GreySource Index[1], Who's Who in Grey Literature[2] , and Conference Proposals issuing from the GL-Series – still other collections, resources, and in-house publications would also deserve future consideration.

The work of defining the metadata for these collections, their subsequent data entry, and additional cross-linking indicated the work that was to be undertaken during the months leading up to GL16. It was anticipated that just as GL15 provided the occasion for the launch of the GreyGuide Repository, GL16 would demonstrate its enhanced function as a web access portal. From the perspective of the GreyGuide, this paper renders an ongoing log, while from the perspective of GreyNet it renders a case study in innovative change in the management of information resources.

Background to Change

Some readers may be familiar with the GL15 conference paper dealing with the launch of the GreyGuide Repository[3] in which a number of defining choices were made during the development phase of the repository.

First of all, the focus would not be on best practices, because in principle there can only be one best practice, while the term good practice better serves to accommodate multiple and varied practices, which speak to the complexity and diversity of grey literature.

Secondly, the repository would not only house published but also proposed good practices creating as it were an open forum for researchers and authors to meet and even collaborate.

And thirdly, it was not expected that the acquisition and initial harvest would produce an abundance of records. The project would be a long term investment, but one that would warrant the wait.

GreyNet's Change in Infrastructure

In early 2014, change came on multiple fronts for GreyNet – one of the project partners in the GreyGuide. And, it is this climate of change that provides the further line of discourse in this paper.

The first front on which change presented itself took place in GreyNet's Infrastructure - originating in the Fall of 2013. GreyNet realized that its growing wealth of content was now housed in three separate OAI compliant repositories:

1. OpenGrey[4], which houses GreyNet's collection of full text conference preprints since 2007 with Inist-CNRS (Nancy, France) as service provider;
2. DANS Easy[5] – Data Archive, which houses GreyNet's research data since 2012 with DANS-KNAW (The Hague, Netherlands) as service provider, and
3. GreyGuide[6] , which in early 2014 came to serve as GreyNet's web access portal with ISTI-CNR (Pisa, Italy) as the service provider.

It became apparent during the test phase of GreyGuide's portal function that a committee should be formed in which GreyNet's three service providers would be better able to collaborate and cooperate in the development not only for GreyNet's resources but also as example or case study for open access to other grey literature web-based resources.
Interestingly, this idea of a committee dealing with GreyNet's open access to resources spread to the other areas of GreyNet's four fold mission to research, publication, open access, and education in the field of grey literature. It became apparent that they too would be better served by the work of committees. And thus GreyNet's new infrastructure came to be, where alongside the annual conference Program Committee, a Research Policy Committee (RPC), a Publishing and Communication Committee (CMC), as well as an Education and Training Committee (LIS) were born .

GreyNet's Resource Policy Committee (RPC)

In October 2013, GreyNet as content provider along with the service providers : Inist-CNRS (France), ISTI-CNR (Italy), and DANS-KNAW (Netherlands) met in Nancy, France. A representative from OpenAIRE[7] (EU) was also invited, because of the interest shown in harvesting the metadata related to GreyNet's collection of full-text conference preprints in OpenGrey. The work of harvesting and integrating this collection was accomplished in December 2014 just prior to GL16 and is itself a demonstration of collaboration actually achieved between and among GreyNet's service providers.
Another example of the earliest collaborative achievement by the Resource Policy Committee (RPC) is the seminar planned in the spring of 2014 in which European organizations involved in the production, open access, and publication of grey literature resources were brought together in what has come to be known simply as the Pisa Seminar.

The Pisa Seminar

Under the direction of colleagues from ISTI-CNR in Pisa, the planning, logistics, and technical support for the Pisa Seminar was organized, while GreyNet set out to identify invited speakers for the program, who would focus on grey literature and policy development related to open access, open data, open standards, and open science. The seminar was scheduled within the GreyForum Series and carried the title 'Grey Literature and Policy Development: The Pisa Declaration'[8] The further task of the RPC was to draft an outline of a document that would fuel discussion and provide a deliverable commensurate to the occasion. That document is entitled "The Pisa Declaration on Policy Development for Grey Literature Resources"[9], which is often simply referred to as the 'Pisa Declaration'.
Over sixty-five participants from prominent European organizations attended the seminar, where an initial draft of the Pisa Declaration was openly discussed – producing constructive comments and textual recommendations.

Publication of the Pisa Declaration


Pisa Declaration
on
Policy Development for Grey Literature Resources

Introduction

A wealth of knowledge and information is produced by organizations, governments and industry, covering a wide range of subject areas and professional fields, not controlled by commercial publishing. These publications, data and other materials known as grey literature, are an essential resource in scholarly communication, research, and policy making for business, industry, professional practice, and civil society. Grey literature is recognized as a key source of evidence, argument, innovation, and understanding in many disciplines including science, engineering, health, social sciences, education, the arts and humanities.

Grey literature document types in print or electronic formats include among others: research and technical reports, briefings and reviews, evaluations, working papers, conference papers, theses, and multimedia content, representing an important and valuable part of research and information. In order to realize the benefits of research and information for scholarship, government, civil society, education and the economy, We, the signatories to this declaration, call for increased recognition of grey literature's role and value by governments, academics and all stakeholders, particulary its importance for open access to research, open science, innovation, evidence-based policy, and knowledge transfer.

May 16, 2014

To achieve the full benefits of grey literature for local, national and global communities we call for and encourage the following:

Organizational

1. Greater commitment to open access by governments and organizations.
2. Greater cooperation and coordination among organizations engaged in the production, use, collection and management of grey literature.
3. The use of persistant identifiers and open metadata standards for grey literature.

Research/Educational

4. New forms of recognition and reward for quality grey literature materials by governments, universities and other institutions.
5. Improved standards in the production and bibliographic control of grey literature.
6. Development and implementation of interoperable standards in the management of grey literature.
7. Development of good practice guides for the production, dissemination, and evaluation of grey literature.

Legal

8. Changes to legal deposit and copyright law that enhance the capacities of libraries, collecting services and educational institutions and programs to collect and provide access to grey literature, particularly non-commercial public interest materials.
9. Addressing legal obstacles to the dissemination of grey literature.
10. Further strides in licensing grey content for both commercial and non-commercial purposes.

Financial/Sustainable

11. Identifying available funding for research involving grey literature.
12. Increased support for collection development and long term preservation of grey literature.
13. Increased investment in infrastructure and new technologies for accessing and using print and digital grey literature.

Technical

14. Strategies to tackle link rot and enhance the stability and accessibility of online content.
15. Systems for linking data and other non-textual content to their grey literature publications together with interoperability standards for sharing grey literature.

To endorse the Pisa Declaration:

Contact

Grey Literature Network Service

Within five weeks after the Pisa Seminar another meeting of the Resource Policy Committee was held at Inist-CNRS in France tasked with incorporating the seminar and post-seminar comments and recommendations into a final draft. The document envelops a 15 point roadmap for grey literature that will guide information professionals well into this 21st Century. The Pisa Declaration was published on May 16th 2014 and immediately appeared on the homepage of the GreyGuide, where it remains for online endorsement by both individuals as well as organizations in the international grey literature community.
An additional task at that Nancy meeting was to exchange standardized metadata developed and used for GreyNet's conference preprints in the OpenGrey repository with the management team at ISTI-CNR responsible for the GreyGuide repository. It is the latter's task to carry out a comparison of the metadata fields that were and would be implemented in

the GreyGuide. Through this exchange, revisions were made in line with the Dublin core standard.

Lobbying the Pisa Declaration

Since its publication, the Pisa Declaration has received some 110 endorsements from information professionals in over twenty countries worldwide. And, it became the objective of the RPC at the GL16 conference to lobby the Pisa Declaration – not only to increase the number of endorsees but also to raise the awareness of the conference participants to the benefits that this declaration could have on policy development for grey literature. In so doing, participants would be able to return home and inform their own professional communities firsthand. The GL16 Poster Session and Sponsor Showcase provided this opportunity and a number of comments were recorded as such:

1. Briefly read through the 15 points that are grouped into one of five categories and focus on one point that is currently most relevant to your workplace;
2. Shake the feeling of going it alone and scroll down the list of names and organizations of those who have already endorsed the document; and
3. Do not hesitate in approaching management with your plans for grey literature, because now you speak the language they understand.

Furthermore, it was during the GL16 Poster Session that a number of non-native English speaking participants voiced their willingness to translate the document into their own language and in some cases in non-roman alphabets. These translations have already begun to be published on the GreyGuide alongside the original English version and will no doubt prompt a new wave of endorsements.

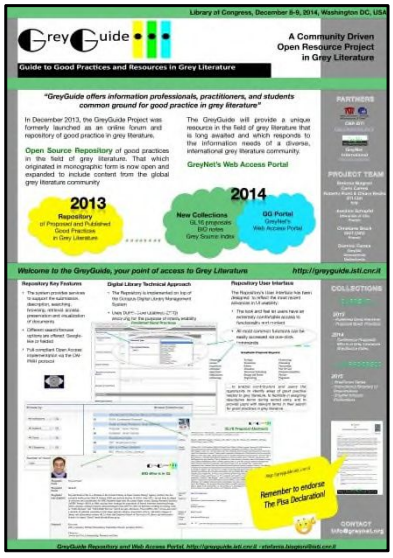
GreyNet's Response to the Pisa Declaration

With the Pisa Declaration on Policy Development for Grey Literature Resources now published, it was GreyNet's turn not only to publicly endorse it but also to implement it. For in this way, the Pisa Declaration could begin to demonstrate the strengths and opportunities grey literature offers, while at the same time exposing the weaknesses and threats facing it. No longer need one be resigned that grey literature is hard to find, but instead how can we best search and access it. No longer hold in question its value and worth, but instead set out the review process it has undergone. And, no longer hesitate as to whether it is published or not, but instead cite and reference grey literature – make it openly public – isn't that what published means?

The Pisa Declaration unveiled an opportunity for GreyNet to migrate its in-house resources that were spread across two websites on multiple webpages in various formats to one web access portal. This would allow for combined browse, search, and retrieval.

Change follows on two fronts

In order to take advantage of the opportunity, change was necessary for both the GreyGuide as well as for GreyNet. In order for it to develop its portal function, the subtitle of the GreyGuide was expanded to read 'Guide to Good Practices and Resources in Grey Literature'. GreyNet on the other hand would have to rethink and revise its workflow in line with the selection and order in which the migration of its resources and collections would take place. The following determinations were made:



- **GSI Collection (GreySource Index)**

The GreySource Index, which was first used to test the planned migration, is a collection of over 70 web based resources in grey literature. This index was originally intended to expose grey literature to the average net-user and in so doing profile organizations responsible for their production and/or processing. Only web-based resources that explicitly refer to the term grey literature (or its equivalent in any language) were listed. GreySource identifies the hyperlink directly embedded in a resource. The migration of content for this collection to the GreyGuide began in the 3rd quarter of 2014 with the first release of its standardized metadata format.

- **BIO Collection (Who is in Grey Literature)**
  WHOIS in the field of Grey Literature is a compilation of over 200 biographical records provided by authors in the International Conference Series on Grey Literature. This online resource is maintained by TextRelease, which currently powers GreyNet. Records in this directory appear in alphabetical order by last name of author and each record contains a current email address. In early October 2014, the metadata record format used in the migration of content to the GreyGuide became operational and authors in the GL-Conference Series are encouraged and will be periodically reminded to complete their own online record entry.

- **GL16 Collection (GL Conference Proposals)**
  Participants who sought to present a paper at GL16 were invited to submit an English abstract between 350-400 words. The abstract is required to deal with the problem/goal, the research method/procedure, an indication of costs related to the project, as well as the anticipated results of the research. The abstract should likewise include the title of the proposed paper, theme(s) most suited to the paper, name(s) of the author(s), and full address information. Since abstracts are the only tangible source that allows the Program Committee to guarantee the content and balance in the conference program, authors are asked to comply with the guidelines. In previous conferences in the GL-Series, these abstracts were accessible on the conference site. However, after a conference took place abstracts were removed from the website. Now however, these conference proposals deposited in a standardized metadata format will remain housed in the GreyGuide.

- **GFS Collection (GreyForum Series)**
  The GreyForum is a series of onsite and online courses, seminars, and workshops where grey literature provides common ground for information professionals in the process of knowledge transfer. The first three events in this series deal with grey literature related to information ethics, information rights, and policy development. It is the documents and documentation produced for these and future events in the GreyForum Series that will be entered in the GreyGuide. The migration of content is scheduled to commence by mid-2015.

- **IDGL Collection (International Directory of Organizations in Grey Literature)**
  The International Directory of Organizations in Grey Literature provides a list of over 200 organizations in 36 countries worldwide that are currently associated with GreyNet either via partnership, membership, sponsorship, or authorship in the field of grey literature. Entries are alphabetical by country and each entry has an embedded link to the corresponding organization's website. GreyNet is proud to serve a global grey literature community and welcomes additions and revisions to this Directory. The migration of content to the GreyGuide is also scheduled to commence in 2015.

- **PUB Collection (GreyNet In-house Publications)**
  GreyNet throughout its more than twenty years in publishing and corporate authorship has a number of serial and non-serial publications that would perhaps be of interest for the GreyGuide. These include a variety of grey literature document types such as newsletters, program books, workbooks, guides, bibliographies, slide presentations, etc. All of which would be made digitally accessible. The migration of GreyNet's PUB Collection will likewise commence during 2015.

Redesigning the Workflow

In the summer of 2014, the GreyGuide management team again met in Pisa to prioritize GreyNet's collections and resources that were to migrate to the GreyGuide. In a concerted effort to engineer the technical and managerial workflow, online templates had still to be designed and tested for each new collection and resource deemed eligible for migration to the GreyGuide.

Since the GL16 conference proposals had already been selected and approved by the Program Committee, it was decided to enter these records directly in the GreyGuide instead of first placing them on the Conference website. Thirty-one conference proposals now remain accessible

in the GreyGuide[10], whereas in previous years these conference proposals would have been removed from the conference site after the event took place.
This change seen in the workflow for GL conference proposals will likewise be adapted for the other resources and collections mentioned above.

Streamlining Data Input

In the fall of 2014, the GreyGuide Management team once again met in Pisa in order to streamline input and better manage change in the repository. A logbook was drafted and became online accessible first between the project workers at ISTI and GreyNet – later to made accessible to the other project workers once the system is thoroughly tested.
Templates for GreyNet's other collections in line for migration to the GreyGuide were drafted and a more thorough review and editorial revision of their corresponding lead texts, field names, and help screens defined the work at hand.
Just as the direct outcome of the earlier management team meeting in Pisa led to the input of the GL16 Conference Proposals in the GreyGuide, the direct outcome of this meeting was the start in migrating records in the 'Who is in Grey Literature' to the BIO Collection in the GreyGuide.[11] With the enhancement of the metadata fields in this template, it is believed that authors will be more inclined to enter online their own records. The added fields now accommodate photographs, curriculum vitae, keywords, links to social media as well as cross-links to selected works by the authors and/or researchers.
On this same occasion, another more informal meeting took place with the head of the D-NET[12] Team at Pisa, who are responsible for the software services for OpenAIRE. During this meeting, earlier problems with harvesting GreyNet's metadata records from the OpenGrey Repository were further discussed and later communicated to the technical department in Nancy. GL16 provided a targeted date in which those problems could be solved and in which OpenAIRE would be able to successfully harvest GreyNet's complete collection of conference preprints.

Reliance on Community Sourcing

With any repository an increase in traffic is dependent on the increase in records that populate it. The premise upon which the GreyGuide was first developed is that community sourcing will be a major asset in the acquisition of content. This premise once again underlies the migration of GreyNet's in-house resources and collections. As expected, the initial phase in the acquisition of records for the repository primarily relies on channels available through the Grey Literature Network Service. Also as expected, populating the repository remains somewhat time consuming and the first harvest has until now not produced an abundance of records. The project however is long term and deemed all the more worthwhile. Based on community sourcing, the GreyGuide will provide a unique resource in the field of grey literature that is long awaited and which responds to the information needs of a diverse, international subject-based community.

Initial Figures and a Traffic Report

The table below indicates in the left hand column the number of records in the GreyGuide corresponding to each collection as of January 2015. These figures indicate that the repository houses 24 good practices in grey literature, while the other 90 records – divided almost evenly over the three collections that until now have migrated to the GreyGuide – account for nearly 80% of the repository's content.

	GUIDE TO GOOD PRACTICES AND RESOURCES IN GREY LITERATURE
4	Proposed Good Practices
20	Published Good Practices
	INTERNATIONAL CONFERENCE SERIES ON GREY LITERATURE
31	GL16 : Conference Proposal
	GREYSOURCE INDEX
29	GSI : GreySource Index
	WHO IS IN GREY LITERATURE
30	BIO : Who is in Grey Literature
114 Metadata Records along with their full-text Documents	

Based on visits to the GreyGuide Repository from April through November 2014, we find:

Month	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Apr 2014	136	419	1376	2381	35.30 MB
May 2014	533	1864	4748	12328	166.11 MB
Jun 2014	592	1695	3845	9370	164.63 MB
Jul 2014	210	1343	2555	4520	67.89 MB
Aug 2014	337	1295	2553	4367	88.48 MB
Sep 2014	444	1840	5296	10018	146.01 MB
Oct 2014	563	1754	6473	12158	248.34 MB
Nov 2014	503	1785	10206	20044	379.04 MB
Total	**3318**	**11995**	**37052**	**75186**	**1.27 GB**

While the above data may be insufficient to draw firm conclusions and base predictions, they do appear to indicate that GreyNet's content stands to gain from a web access portal and the GreyGuide stands to gain by an increase the number of records housed.

Some Concluding Remarks

This paper speaks of change not only for a subject based repository and an organization that has served the grey literature community for over two decades but also for the entire field of grey literature re-enkindled by the Pisa Declaration. The change described in this paper is not only technology driven but is also community driven. When the opportunity for change presents itself, it may not always be confined to one information product or service, but may open up on multiple fronts calling for an innovative response. One that will not be adequately addressed simply by making improvements, but by doing things differently, by applying better solutions that meet new requirements and the needs of existing information markets. Throughout this paper specific reference is made to dates and places where face to face meetings were held and where work in teams took place. Without that hands-on approach, the change spoken of in this paper and the change that lies ahead could not be implemented.

References

[1] GreySource, A Selection of Web-based Resources in Grey Literature
http://www.greynet.org/greysourceindex.html

[2] WHOIS in the field of Grey Literature
http://www.textrelease.com/whoisbionotes.html

[3] GreyGuide - Guide to Good Practice in Grey Literature: A Community Driven Open Resource Project
http://hdl.handle.net/10068/1024609

[4] OpenGrey - System for Information on Grey Literature in Europe
http://www.opengrey.eu

[5] EASY is the online archiving system of Data Archiving and Networked Services
https://easy.dans.knaw.nl/ui/home

[6] GreyGuide – Guide to good practices and resources in grey literature
http://greyguide.isti.cnr.it/

[7] OpenAIRE - A network of Open Access repositories, archives and journals that support Open Access policies
https://www.openaire.eu/

[8] Grey Literature and Policy Development: The Pisa Declaration http://www.greynet.org/greyforumseries/policydevelopment.html

[9] Pisa Declaration on Policy Development for Grey Literature Resources
http://greyguide.isti.cnr.it/include/pisa-declaration-poster-lastversion3.pd

[10] GL16 Conference Proposals - International Conference Series on Grey Literature
http://greyguide.isti.cnr.it/listtitoli.php?authority=GLConference&collection=GL16&langver=en&RighePag=100

[11] BIO Collection – Who is in Grey Literature
http://greyguide.isti.cnr.it/listtitoli.php?authority=WhoIs&collection=BIO&langver=en&RighePag=100

[12] D-NET: Building Sustainable Aggregative Data Infrastructures
http://www.d-net.research-infrastructures.eu/node/4

Pisa Declaration on Policy Development for Grey Literature Resources

Introduction

A wealth of knowledge and information is produced by organizations, governments and industry, covering a wide range of subject areas and professional fields, not controlled by commercial publishing. These publications, data and other materials known as grey literature, are an essential resource in scholarly communication, research, and policy making for business, industry, professional practice, and civil society. Grey literature is recognized as a key source of evidence, argument, innovation, and understanding in many disciplines including science, engineering, health, social sciences, education, the arts and humanities.

Grey literature document types in print or electronic formats include among others: research and technical reports, briefings and reviews, evaluations, working papers, conference papers, theses, and multimedia content, representing an important and valuable part of research and information.
In order to realize the benefits of research and information for scholarship, government, civil society, education and the economy, We, the signatories to this declaration, call for increased recognition of grey literature's role and value by governments, academics and all stakeholders, particularly its importance for open access to research, open science, innovation, evidence-based policy, and knowledge transfer.

May 16, 2014

To achieve the full benefits of grey literature for local, national and global communities we call for and encourage the following:

Organizational

1. Greater commitment to open access by governments and organizations.
2. Greater cooperation and coordination among organizations engaged in the production, use, collection and management of grey literature.
3. The use of persistent identifiers and open metadata standards for grey literature.

Research/Educational

4. New forms of recognition and reward for quality grey literature materials by governments, universities and other institutions.
5. Improved standards in the production and bibliographic control of grey literature.
6. Development and implementation of interoperable standards in the management of grey literature.
7. Development of good practice guides for the production, dissemination, and evaluation of grey literature.

Legal

8. Changes to legal deposit and copyright law that enhance the capacities of libraries, collecting services and educational institutions and programs to collect and provide access to grey literature, particularly non-commercial public interest materials.
9. Addressing legal obstacles to the dissemination of grey literature.
10. Further strides in licensing grey content for both commercial and non-commercial purposes.

Financial/Sustainable

11. Identifying available funding for research involving grey literature.
12. Increased support for collection development and long term preservation of grey literature.
13. Increased investment in infrastructure and new technologies for accessing and using print and digital grey literature.

Technical

14. Strategies to tackle link rot and enhance the stability and accessibility of online content.
15. Systems for linking data and other non-textual content to their grey literature publications together with interoperability standards for sharing grey literature.

GreyGuide

To endorse the Pisa Declaration:

http://greyguide.isti.cnr.it

Contact
info@greynet.org

GreyNet
Grey Literature Network Service

List of Participating Organizations

African and Middle Eastern Division; Library of Congress	United States
American Psychological Association, APA	United States
Aerospace Corporation, The	United States
Biblioteca Centrale, G. Marconi; CNR	Italy
Canadian Agency for Drugs and Technologies in Health, CADTH	Canada
Centre National de Recherché Scientifique, CNRS	France
Dalhousie University, School of Information Management	Canada
Data Archiving and Networked Services, DANS	Netherlands
Delta Progetti 2000 S.r.l.; COMDATA	Italy
EBSCO Publishing	United States
ETT Solutions Ltd.	Italy
Federal Library Information Network, FEDLINK	United States
Federal Reserve Bank of St. Louis	United States
GERiiCO laboratory	France
Grey Literature Network Service, GreyNet International	Netherlands
Indiana University School of Informatics and Computing, IUPUI	United States
Information International Associates, IIa	United States
Institut de l'Information Scientifique et Technique, Inist	France
Institute of Computational Linguistics, ILC-CNR	Italy
Institute of Information Science and Technologies, ISTI	Italy
Institute for Scientific Networking, ISN Oldenburg	Germany
Istituto Nazionale di Astrofisica, INAF; Osservatorio Astronomico di Bologna	Italy
Istituto Superiore di Sanità, ISS	Italy
Japan Atomic Energy Agency, JAEA	Japan
King Abdullah University of Science and Technology, KAUST	Saudi Arabia
Knowledge Resource Service, KRS	Canada
Korea Institute of Science & Technology Information, KISTI	Korea
Library of Congress, LoC	United States
National Library of Technology, NTK	Czech Republic
National Research Council, CNR	Italy
National Union Catalog of Manuscript Collections, NUCMC	United States
New York Academy of Medicine, NYAM	United States
Nuclear Information Section; International Atomic Energy Agency, NIS-IAEA	Austria
Pratt Institute, School of Information and Library Science	United States
PricewaterhouseCoopers, PwC	Netherlands
Research Institute of Geodesy, Topography and Cartography, VUGTK	Czech Republic
Roy Rosenzweig Center for History and New Media, George Mason University	United States
Slovak Centre of Scientific and Technical Information, CVTISR	Slovakia
Slovak University of Technology in Bratislava, STUBA	Slovakia
Swinburne University of Technology, SWIN	Australia
TextRelease, Program and Conference Bureau	Netherlands
Tisch Library, Tufts University	United States
Université Charles de Gaulle - Lille 3	France
University Hospitals Bristol, BRI; NHS Foundation Trust	United Kingdom
University of Calgary	Canada
University of California, Berkeley School of Information	United States
University of California, Irvine Libraries, UCI	United States
University of Lille 3	France
Uppsala University	Sweden
University of Wisconsin-Milwaukee, School of Information Studies	United States
U.S. Department of Justice, Civil and Criminal Division	United States
U.S. Department of Transportation, DOT	United States

Aloia, Danielle **106**
Danielle Aloia received her MSLS from Catholic University of America, in Washington, DC, in 2005 while working on the AgeLine Database at AARP. She has over 20 years of experience in a variety of library settings, including academic, non-profit and museum. She has been involved with collecting, evaluating, and cataloging grey literature beginning in 2006 at AARP and then at the United States. Dept. of Transportation. For the past 3 years she has been managing the Grey Literature Report in Public Health, produced by the New York Academy of Medicine. Email: daloia@nyam.org

Berchum, Marnix van **133**
Marnix van Berchum is Head of Data Services at DANS, responsible for the coordination of the services DANS offers (including EASY, the narcis.nl portal, and the Dutch Dataverse Network). After his studies in Musicology at Utrecht University, he started working at Utrecht University Library. In the years 2010-2012 he also worked at SURF. At both employers he was mostly involved with projects related to Open Access and innovations in scholarly communications (e.g. OpenAIRE). Marnix combines his work at DANS with a PhD trajectory at Utrecht University, in which he wants to apply the concepts and methods of network theory on the dissemination of music in the 16th century. Marnix is Associate Director of the CMME Project. Email: marnix.van.berchum@dans.knaw.nl

Biagioni, Stefania **17, 147**
Stefania Biagioni graduated in Italian Language and Literature at the University of Pisa and specialized in data processing. She is currently a member of the research staff at the Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI), an institute of the Italian National Research Council (CNR) located in Pisa. She is head librarian and member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She has been the responsible of ERCIM Technical Reference Digital Library (ETRDL) and currently of the PUMA (PUblication MAnagement) & MetaPub, a service oriented and user focused infrastructure for institutional and thematic Open Access repositories looking at the DRIVER vision, http://puma.isti.cnr.it. She has coauthored a number of publications dealing with digital libraries. Her activities include integration of grey literature into library collections and web access to the library's digital resources, including electronic journals and databases. She is a member of GreyNet since 2005. Email: stefania.biagioni@isti.cnr.it

Börjesson, Lisa **93**
Lisa Börjesson is a doctoral student at the Department of archival science, library & information science, and museum & heritage studies at Uppsala University, Sweden. She also holds a position as a visiting student researcher at UC Berkeley School of Information 2014-2015. Lisa has a background as a library development consultant working within regional government. She currently works together with archaeologists and information scholars in the research project Archaeological Information in the Digital Society (ARKDIS) funded by the Swedish Research Council, and is a member of the northwestern European Centre for Digital Heritage. Her ongoing dissertation research (2013-2017) explores three contextual factors shaping reporting in professional archaeology: information policy, report writers' frames of reference and their interpretations of the report writing as a professional task. Her research complements initiatives to improve technical access to archaeological grey literature by providing a more nuanced understanding for the report genre. Email: lisa.borjesson@abm.uu.se

Choi, Seonheui **101**
Seonheui Chio is the manager of dept. of domestic information at the Korea Institute of Science and Technology Information (KISTI). She is a one of OAK project manager, phase 2. Her research interests include: Open Access journal publishing, DOI, Institutional Repositories.
Email: sunny.choi@kisti.re.kr

Copeland, Andrea J. **41**
Andrea J. Copeland is an assistant professor in the School of Library and Information Science at Indiana University at Indianapolis. She received a PhD from Drexel University in information studies in September 2009. The digital preservation practice of public library users is the subject of her dissertation. Her studies at Drexel were supported by the Laura Bush 21st Century Librarian Fellowship from the Institute of Museum and Library Services. She was also the recipient of the Society of Woman Geographers Fellowship in 2002 for her research on the variation of public library use in New York City neighborhoods. Email: ajapzon@iupui.edu

Crowe, June **11**
June Crowe is the Senior Researcher and Group Manager, Open Source Research Division at Information International Associates, Inc. (IIa). She received her AMLS from the University of Michigan, Ann Arbor, and her M.Ed. in geographic education from the University of Georgia, Athens. Her current position, subject matter expert cultural studies, is with the Homeland Defense and Security Information Analysis Center (HDIAC) in Oak Ridge, TN. Her primary interests are Grey Literature research, socio-cultural and S&T research, digital repositories, and open source intelligence tools. Email: jcrowe@iiaweb.com

Drozda, Jiří **59**
Jiří Drozda is currently chief of Information Department and Surveyors Library of Czech Research Institute of Geodesy, Topography and Cartography (VUGTK). He graduated in Geodesy and Cartography in 1984 at Military University in Brno (CR), special courses in 1997 at Defense Mapping School in Fort Belvoir (USA) and in 2008 at NATO Communication and Information Systems School in Latina (IT). Before coming to VUGTK, he worked for Military Geographic Institute as deputy director (1996-2003) and for NATO as International Civilian Consultant for geographic information support in Balkan territory. He is involved in various national and EU projects focused to education and dissemination of geographic information. Email: jiri.drozda@vugrk.cz

Farace, Dominic **147**
Dominic Farace is Head of GreyNet International and Director of TextRelease, an independent information bureau specializing in grey literature and networked information. He holds degrees in sociology from Creighton University (BA) and the University of New Orleans (MA), and has a doctorate in social sciences from the University of Utrecht, Netherlands – where he lives and works since 1976. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (1986-1992), Farace founded GreyNet, Grey Literature Network Service. He has since been responsible for the International Conference Series on Grey Literature (1993-2015). In this capacity, he also serves as Program and Conference Director as well as managing editor of the Conference Proceedings. He is editor of The Grey Journal and provides workshops and training in the field of grey literature.
Email: dominic.farace@textrelease.com

Gelfand, Julia M. **120**
Julia Gelfand has served as Applied Sciences and Engineering Librarian at the University of California, Irvine Libraries since 1981, where she was recognized as a Distinguished Librarian. She has presented and written extensively on various issues related to grey literature, collection development, scholarly communication, library vendor relationships, emerging technologies, scholarly publishing, social media, and organizational change in academic libraries. She is active in many professional organizations and currently serves on the Board of Directors of the Association of College and Research Libraries (ACRL) Division of the American Library Association (ALA), as Secretary of Section T for Information, Computers & Communication of the American Association for the Advancement of Science (AAAS), and is a member of the Acquisitions & Collection Development Section of the International Federation of Library Associations (IFLA). Email: jgelfand@uci.edu

Giannini, Silvia **17**
Silvia Giannini graduated and specialized in library sciences. Since 1987 she has been working in Pisa at the Institute for the Science and Technologies of Information "A. Faedo" of the Italian National Council of Research (ISTI-CNR) as a librarian. She is a member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She is responsible of the library automation software "Libero" in use at the CNR Research Area in Pisa and coordinates the bibliographic and managing activities of the ISTI library team. She cooperates in the design and development of the PUMA (PUblication MAnagement) & MetaPub, an infrastructure software for institutional and thematic Open Access repositories of published and grey literature produced by CNR. Email: silvia.giannini@isti.cnr.it

Goggi, Sara **17, 108**
Sara Goggi is a technologist at CNR-ILC. She started working at ILC in 1996 working on the EC project LE-PAROLE for creating the Italian reference corpus; afterwards she began dealing with the management of several European projects and nowadays she is involved with organisational and managerial activities mainly concerning international relationships and dissemination as well as organization of events (e.g. LREC conference series). Currently one of her permanent activities is the editorial work for the international ISI Journal Language Resources and Evaluation, being its Assistant Editor. Since many years (from 2004) she also carries on research on terminology: this field of research starting intersecting with work on Grey Literature in 2011 at GL13. Email: sara.goggi@ilc.cnr.it

Gonda, Mayuki **139**
Mayuki Gonda works as librarian at the Central Library of JAEA (Japan Atomic Energy Agency). He joined JAERI (former JAEA) in 2005, and had been working for management and dissemination of JAEA research results information. Since 2009, he is in charge of selection, classification and indexing for INIS (International Nuclear Information System) Database. He is also a member of editorial committee of the Journal of Information Science and Technology Association (INFOSTA) since 2008. He holds a degree in information science (M.A.) from the Graduate School of Library, Information and Media Studies, University of Tsukuba. Email: gonda.mayuki@jaea.go.jp

Hayakawa, Misa **139**
Misa Hayakawa works as librarian at the Central Library of Japan Atomic Energy Agency(JAEA). She Joined JAEA in 2010, and has been working for managing metadata of both papers and oral presentations published by JAEA researchers. In addition, she disseminates such information on the Internet via "JAEA Originated Papers Searching System"(JOPSS). She got the master's degree in Library and Information Science at the Graduate School of Library Information and Media Studies, University of Tsukuba in Japan. hayakawa.misa@jaea.go.jp

Hwang, Hyekyong **101**
Hyekyong Hwang is a senior researcher at the Korea Institute of Science and Technology Information (KISTI). She has Ph.D in library and information science from Yonsei University (2014). From 2010 until 2014 she has been participated in OAK(Open Access Korea) project in the areas of institutional repositories. Her research interests include: Open Access, Adoption of Institutional Repositories, User Behavior, and Grey Literature. Email: hkhwang@kisti.re.kr

Ikeda, Kiyoshi **139**
Kiyoshi Ikeda works as a librarian at the Central Library of Japan Atomic Energy Agency (JAEA). He also currently works as the chief of International Nuclear Information System (INIS) and Nuclear Information Management Section of JAEA. Since 2011, he has been in charge of selection and classification for INIS Database input data. He joined Japan Atomic Energy Research Institute (JAERI, former JAEA) in 2004, and has been working for acquisition of library materials (domestic/foreign books, journals, technical reports and proceedings). From 2007 to 2010, he had been working for editing work at "JAEA Reports" and "JAEA R&D Review" which are grey literatures. From 2008 to 2013, he has also been a member of editorial committee of the Japan Special Libraries Association (JSLA). He holds a degree in library and information science (M.A.) from the Graduate School of University of Library and Information Science (in Japan).
Email: ikeda.kiyoshi@jaea.go.jp

Kim, Hyesun **101**
Hyesun Kim is the manager of dept. of NDSL Service at the Korea Institute of Science and Technology Information (KISTI). She has Ph.D. in Library and Information Science from Ewha Women's University (2012). She had leading the OAK(Open Access Korea) project, phase 1. Her research interests include: information services, collection development for journals, Open Access. Email: hskim@kisti.re.kr

Lipinski, Tomas A. **41**
Tomas A. Lipinski obtained his J.D. from Marquette University Law School, LL.M. from The John Marshall Law School, and Ph.D. from the University of Illinois at Urbana-Champaign. Professor Lipinski has worked in a variety of library and legal settings including the private, public and non-profit sectors. Professor Lipinski teaches researches and speaks frequently on various topics within the areas of information law and policy, especially copyright, free speech and privacy issues in schools and libraries. In fall of 2005, Professor Lipinski was placed on the Fulbright Senior Specialist Roster and was named a member of the Global Law Faculty, University of Leuven in Fall of 2006. Email: lipinski@sois.uwm.edu

MacDonald, Bertrum H. **31**
Bertrum H. MacDonald is a Professor of Information Management in the School of Information Management at Dalhousie University, Halifax, Canada. With a background in science (BSc, Biology), history of science (MA), and information science (MLS, PhD), he pursues research that investigates the dissemination and use of scientific information in historical and contemporary contexts. He pursues interdisciplinary research, particularly within the Environmental Information: Use and Influence initiative (www.eiui.ca), since this work tackles large questions from the point of view of several relevant disciplines. He has been Director of the School of Information Management and Associate Dean (Research) in the Faculty of Management at Dalhousie University. He can be seen speaking about research projects at local, national, and international levels, and he holds executive positions with national and international associations. In 2004, he won the International GreyNet Award with his research colleagues, Ruth Cordes and Peter Wells. He is the recipient of the Marie Tremaine Medal, the highest award of the Bibliographical Society of Canada, and he was awarded a Dibner Research Fellowship at the Smithsonian Institution in 2001. Email: bertrum.macdonald@dal.ca

Pardelli, Gabriella **17, 108**
Gabriella Pardelli was born at Pisa, graduated in Arts in 1980 at the Pisa University, submitting a thesis on the History of Science. Since 1984, researcher at the National Research Council, Institute of Computational Linguistics "Antonio Zampolli" ILC, in Pisa. Head of the Library of the ILC Institute since 1990, responsible for the Archives of the ILC Institute since 2005. Her interests and activity range from studies in grey literature and terminology, with particular regard to the Computational Linguistics and its related disciplines, to the creation of documentary resources for digital libraries in the Humanities. She has participated in many national and international projects including the recent projects:- BIBLOS: Historical, Philosophical and Philological Digital Library of the Italian National Research Council, (funded by CNR); - For digital edition of manuscripts of Ferdinand de Saussure (Research Programs of Relevant National Interest, PRIN - funded by the Ministry of Education, University and Research, MIUR). Email: gabriella.pardelli@ilc.cnr.it

Pejšová, Petra **59, 65**
Petra Pejšová graduated in Information science and librarianship at the Charles University in Prague. Currently she is a manager of the National repository of Grey Literature in the Czech Republic. Its goal is a systematic collection, long-term archival and provision of access to specialized grey literature, especially to do with research and development, civil service and education, as well as from the business sphere and "open access" at the national level. She works in the National Library of Technology as a head of the Digital National Technical Library. Through lectures and publications she provides information on grey literature in the Czech Republic. She is a member of Czech affiliate team for dissemination free licenses Creative Commons. Email: petra.pejsova@techlib.cz

Prost, Hélène **75**
Hélène Prost is information professional at the Institute of Scientific and Technical Information (CNRS) and associate member of the GERiiCO research laboratory (University of Lille 3). She is interested in empirical library and information sciences and statistical data analysis. She participates in research projects on evaluation of collections, document delivery, usage analysis, grey literature and open access, and she is author of several publications. Email: prost@inist.fr

Schöpfel, Joachim **75**
Joachim Schöpfel is Lecturer of Library and Information Sciences at the University of Lille 3 (France), Director of the French Digitization Centre for PhD theses (ANRT) and member of the GERiiCO research laboratory. He was Manager of the INIST (CNRS) scientific library from 1999 to 2008. He teaches Library Marketing, Auditing, Intellectual Property and Information Science. His research interests are scientific information and communication, especially Open Access and Grey Literature.
Email: joachim.schopfel@univ-lille3.fr

Sherline, Crystal **11**
Crystal Sherline is a Technical Information Analyst for Homeland Defense and Security Information Analysis Center (HDIAC), a researcher at Information International Associates (IIA), and doctoral candidate at the University of Tennessee, Knoxville. She received a Master's degree from the University of Maryland, and an MLIS from Florida State University, and will be receiving her Ph.D. from the University of Tennessee in July 2014. She has experience in operations of both public and academic libraries. Her interests are varied, but consist of primarily research, including Network Analysis, Grey Literature, HCI, Social Science, and data sharing. Email: csherline@iiaweb.com

Soomai, Suzuette S. **31**
Suzuette S. Soomai is an interdisciplinary doctoral student in the Faculty of Graduate Studies. Her research focuses on the role of fisheries information, published extensively by national and international governmental organisations, in policy making for fisheries management. She holds a Master of Marine Management from Dalhousie University and an MPhil (Zoology/Aquatic Ecology) and BSc from the University of the West Indies, Trinidad and Tobago. Prior to her studies at Dalhousie University, Suzuette was a Fisheries Officer with the Ministry of Agriculture, Land and Marine Resources in Trinidad and Tobago and was a member of fisheries scientific working groups led by the United Nations Food and Agriculture Organisation and the Caribbean Regional Fisheries Mechanism. She has special expertise in fisheries resource and coastal zone management and has completed national and regional fish stock assessments while interacting with a diverse range of stakeholders in the Caribbean. In 2012, she was awarded a major doctoral fellowship from the Social Sciences and Humanities Research Council of Canada. Email: suzuette.soomai@dal.ca

Stock, Christiane **147**
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 OpenGrey. 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

Synková, Veronika **59**
Veronika Synková studied information science and librarianship at the Institute of Information Studies and Librarianship at the Faculty of Arts of Charles University in Prague, Czech Republic. She works as a system librarian in the Library of Antonin Svehla of the Institute of Agricultural Economics and Information. She is a member of the team responsible for the library system Aleph and the discovery system Primo. She cooperates on implementation of the system Invenio with the Surveying Library of the Research Institute of Geodesy, Topography and Cartography.
Email: veronika.synkova@gmail.com

Tsang, Daniel C. **120**
Daniel C. Tsang has served as Social Sciences Bibliographer and Data Librarian at the University of California, Irvine Libraries since 1986 and was named Distinguished Librarian. He is a well published author in mainstream and alternative media and invited speaker concentrating on anti-surveillance and social movements. Recognized for his scholarship and contributions to the documentation of small and alternative presses, and for his distinguished service to data librarianship, he was awarded the William F. Flanigan Award for Distinguished Service to ICPSR in 2013. He has graduate degrees from the University of Michigan. Email: dtsang@uci.edu

Vaska, Marcus **65**
Marcus Vaska is a librarian with the Knowledge Resource Service (KRS), University of Calgary, responsible for providing research and information support to staff affiliated with an Alberta Cancer Care research facility. A firm supporter of embedded librarianship, Marcus engages himself in numerous activities, including instruction and research consultation, with numerous research teams. An advocate of the Open Access Movement, Marcus' current interests focus on creating greater awareness of the importance of creative commons licenses in the publishing of grey literature material. Email: mmvaska@ucalgary.ca

Wells, Peter G. **31**
Peter G. Wells is an Adjunct Professor in the School for Resource and Environmental Studies and the Marine Affairs Program, Faculty of Management, and a Senior Research Fellow, International Ocean Institute, at Dalhousie University, Halifax, Canada. He holds a BSc (Biology) from McGill University, a MSc (Zoology) from the University of Toronto, and a PhD (Zoology/Aquatic Toxicology) from the University of Guelph. After over 30 years of public service, he took early retirement from Environment Canada, in 2006, to focus on his research. He concluded his work with Environment Canada as Head, Coastal and Water Science, and Senior Research Scientist, Coastal Ecosystems. He served the United Nations Joint Group of Scientific Experts on Marine Environmental Protection in various capacities for 14 years, and taught an international marine pollution course in Bermuda for 16 years. His current research includes choosing indicators for coastal ecosystem health, utilizing blue mussels for monitoring chemical contaminants, and evaluating the use and influence of marine environmental information in environmental policies and decision making. He was elected Fellow of the American Association for the Advancement of Science, and is a recipient of Dalhousie's highest award for teaching excellence by part-time faculty.
Email: oceans2@ns.sympatico.ca

Seventeenth International Conference on Grey Literature

'A New Wave of Textual and Non-Textual Grey Literature'

Royal Netherlands Academy of Arts and Sciences
Amsterdam, Netherlands on December 1-2, 2015

Announcement

As the internet becomes increasingly grey and every cloud now has a grey lining, there arises the need to address a new and challenging wave of textual and non-textual grey literature. GL17 will examine a number of new types of textual grey literature both web-based and submerged in the sea of social networks. No less attention will be drawn to the expanding quantity of non-textual grey literature accessible in visual, audio, and diverse data formats and frequencies.

Actually, in order to grasp this new wave of grey literature it may be even more advantageous to look at the convergence of these new types of textual and non-textual content rather than focus separately on each. The problems textual grey literature faced and addressed over the past quarter century are to a certain extent very similar to what non-textual grey literature faces today. The wide range of graphics, photographs, and other data-intensive grey literature is obscure, hard to find, and often short lived because it lacks proper indexing and sustained access. Such non-textual grey literature requires interpretation and documentation, which can in part be achieved by linking and crosslinking to their related textual counterparts. In this way, grey literature becomes leveraged and its value and return on investment made transparent.

While bridging textual and non-textual content is technically possible, it also requires an information policy in place that supports these new digital assets. Likewise, information professionals and practitioners must be able to (re)appropriate human resources and streamline their workflow in innovative ways. These should allow for content and feedback generated in social networks and in particular the information communities served.

The Seventeenth International Conference on Grey Literature welcomes authors and researchers to share their experience and vision on how to channel this new wave of grey literature by responding to the GL17 Call for Papers, http://www.textrelease.com/gl17callforpapers.html.

Conference Related Topics

New Types of Grey Literature Documents
New Forms of Non-Textual Grey Literature
The Convergence of Textual and Non-Textual Grey Literature
The Influence of Social Networks on Grey Literature
Innovative ways in Leveraging Grey Literature

NOTE To keep conference delegates and participants informed on the progress of the Seventeenth International Conference on Grey Literature, a quarterly newsletter, conference calendar, and further documentation will be made available on the conference site, www.textrelease.com. Please direct all correspondence to conference@textrelease.com.

TextRelease
GL17 Program and Conference Bureau

Javastraat 194-HS, 1095 CP **A**msterdam, **T**he **N**etherlands
www.textrelease.com · conference@textrelease.com
Tel/Fax +31-20-331.2420

Seventeenth International Conference on Grey Literature

'A New Wave of Textual and Non-Textual Grey Literature'

Royal Netherlands Academy of Arts and Sciences
Amsterdam, Netherlands on December 1-2, 2015

Call for Papers

Title of Paper:	**Related Topic(s):**
Author Name(s):	**Telephone:**
Organization(s):	**Fax:**
Postal Address:	**Email:**
Post/Zip Code - City - Country:	**URL:**

Guidelines for Abstracts

Participants who seek to present a paper at GL17 are invited to submit an English abstract between 350-400 words. The abstract should deal with the problem/goal, the research method/procedure, an indication of costs related to the project, as well as the anticipated results of the research. The abstract should likewise include the title of the proposed paper, topic(s) most suited to the paper, name(s) of the author(s), and full address information. Abstracts are the only tangible source that allows the Program Committee to guarantee the content and balance in the conference program. Every effort should be made to reflect the content of your work in the abstract submitted. Abstracts not in compliance with the guidelines will be returned to the author for revision.

Conference Related Topics

- ☐ **New Types of Grey Literature Documents**
- ☐ **New Forms of Non-Textual Grey Literature**
- ☐ **The Convergence of Textual and Non-Textual Grey Literature**
- ☐ **The Influence of Social Networks on Grey Literature**
- ☐ **Innovative ways in Leveraging Grey Literature**
- ☐ **Other Conference Topic**

Due Date and Format for Submission

Abstracts in MS Word must be emailed to conference@textrelease.com on or before **April 24th 2015**. The author will receive verification upon its receipt. Shortly after the Program Committee meets in mid-May, authors will be notified of their place on the conference program. This notice will be accompanied by further guidelines for submission of full text papers, accompanying research data, PowerPoint slides, and required Author Registration.

TextRelease
GL17 Program and Conference Bureau

Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com · conference@textrelease.com
Tel/Fax +31-20-331.2420

Forthcoming
March 2015

Grey Literature Lobby: Engines and Requesters for Change

The Library of Congress, Washington D.C. USA on 8-9 December 2014

Publication Order Form

SIXTEENTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

Publication(s):	*No. of Copies*	*x*	*Amount in Euros*	Subtotal
GL16 Conference Proceedings - Printed Edition ISBN 978-90-77484-23-4 ISSN 1386-2316 *Postage and Handling* ***excluded****)		x	109.00 =	€
GL16 Conference Proceedings - CD-Rom Edition ISBN 978-90-77484-23-4 ISSN 1386-2316 *Postage and Handling* ***included***		x	109.00 =	€
GL16 Conference Proceedings - Online Edition ISBN 978-90-77484-23-4 ISSN 2211-7199 ***Password Protected Access***		x	109.00 =	€

*POSTAGE AND HANDLING PER PRINTED COPY*****)**

Holland		x	5.00	
Europe		x	10.00	€
Other		x	15.00	€
			TOTAL EURO =	**€**

Customer Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
E-mail Address:	

Upon receipt of payment the publication(s) will be forwarded to your shipping address with an invoice marked paid.

❑ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to "GL16 Publication Order"

❑ MasterCard/Eurocard ❑ Visa Card ❑ American Express

Card No. _____ Expiration Date: _____

Print the name that appears on the credit card, here _____

Signature: _____ **CVC II code:** _____ *(Last 3 digits on signature side of card)*

Place: _____ Date: _____

NOTE: CREDIT CARD TRANSACTIONS CAN BE AUTHORIZED BY PHONE, FAX, OR POSTAL SERVICES. EMAIL IS NOT AUTHORIZED.

TextRelease
www.textrelease.com

GL16 Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: info@textrelease.com

Index to Authors