Public Interest in Accessing the INIS Collection

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

Abstract
Since its creation in 1970, the International Nuclear Information System (INIS) has collected and provided access to more than 3.8 million bibliographic references to publications, documents, technical reports, non-copyrighted documentation, and other grey literature, as well as over a million full texts. Overall, there are 800 GB of data in the INIS repository. Public interest in accessing this collection has been remarkable. This year alone, there were more than one million sessions and almost two million page views. During the same period, there were 1.6 million full text document downloads.
The INIS collection consists of seven types of literature – computer media, patents, books, reports, journal articles, miscellaneous and audio-visuals. This paper provides an overview of the INIS collection subject coverage, and the distribution of different types of grey literature. It also provides INIS repository access statistics based on Google Analytics and other Web search data pertaining to public interest in accessing these different types of literature and the uniqueness of the collection.
As one of the world's largest collections of published information on the peaceful uses of nuclear science and technology, INIS represents an extraordinary example of world cooperation. 154 INIS members share and allow access to their valuable nuclear information resources, preserving them for future generations and offering a freely available nuclear knowledge repository.

Keywords: Nuclear information; Document repository; Grey literature; INIS; IAEA

Introduction

Since its creation in 1970, the International Nuclear Information System (INIS) has collected and provided access to more than 3.8 million bibliographic references to publications, documents, technical reports, non-copyrighted documentation, and other grey literature, as well as over a million full texts. The INIS repository hosts one of the world's largest collections of published documentation and information on the peaceful uses of nuclear science and technology.

Figure 1: Beginning of automation in INIS

This paper reviews the public interest in accessing the INIS collection. Although who accesses INIS is an important question, greater emphasis is placed here on the question of why there is public interest in accessing this valuable collection of nuclear grey literature.
It is assumed that the reasons for public interest lie in three main features of the collection, namely, its subject coverage, type of documents and uniqueness. These three features are the focus of this paper. Available website analytics are used to confirm these assumptions and, at the same time, offer possible direction for further improvements in INIS collection services.

International Atomic Energy Agency (IAEA)

The IAEA is regarded as the world's centre of cooperation in the field of safe, secure and peaceful uses of nuclear technologies. It was set up in 1957 as the world´s "Atoms for Peace" (IAEA, 2015) organization within the United Nations system. As of November 2015, the IAEA has 166 Member States.

Figure 2: Vienna International Centre

The IAEA Secretariat is headquartered at the Vienna International Centre in Vienna, Austria. It also operates liaison and regional offices in Geneva, New York, Toronto, and Tokyo. The IAEA runs and supports research centres and scientific laboratories in Vienna and Seibersdorf, Austria; Monaco; and Trieste, Italy. The IAEA Secretariat is a team of 2560 multidisciplinary professional and support staff from more than 100 countries.

The IAEA's mission is guided by the interests and needs of Member States, strategic plans and the vision embodied in the IAEA Statute. Three main pillars — or areas of work — underpin the IAEA's mission: Safety and Security; Science and Technology; and Safeguards and Verification.
The work of the IAEA is carried out through six departments (IAEA, 2015): Nuclear Energy, Nuclear Safety and Security, Nuclear Science and Applications, Safeguards, Technical Cooperation, and the Department of Management. Although supporting the entire Agency, the Nuclear Information Section (NIS) is organizationally part of the Department of Nuclear Energy. The Department's main tasks are to foster the efficient and safe use of nuclear power by supporting interested Member States in improving the performance of nuclear power plants, the nuclear fuel cycle, and the management of nuclear wastes; catalysing innovation in nuclear power and fuel cycle technologies; development of indigenous capabilities for national energy planning; the deployment of new nuclear power plants; and the advancement of science and industry through improved operation of research reactors.

IAEA Nuclear Information Goals

The main nuclear information goals of the IAEA (IAEA, 1956) are to (a) Foster the exchange of scientific and technical information on the peaceful use of nuclear science and technology, which involves collection, processing, preservation and dissemination of information; (b) Increase awareness of the importance of managing nuclear information resources; (c) Assist with capacity building and training; and (d) Provide information services and support to Member States.
The collection, preservation and dissemination of nuclear information and knowledge is, in turn, the main responsibility of the Nuclear Information Session (NIS), more specifically of INIS.

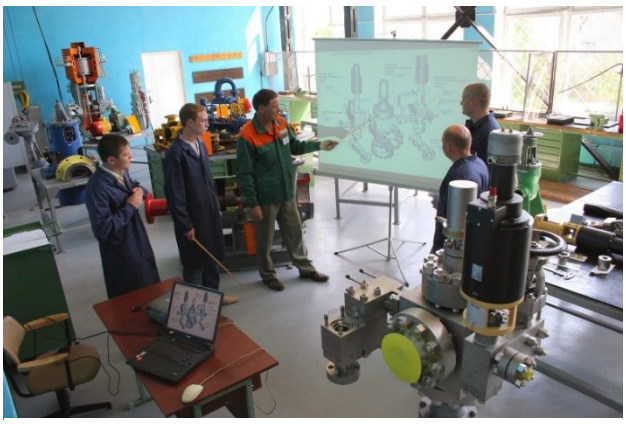


Figure 3: IAEA workshop

NIS consists of the IAEA Library Unit, the INIS Unit and the Systems Development and Support Group. It fosters the exchange of scientific and technical information on the peaceful use of nuclear science and technology; increases awareness in Member States of the importance of maintaining efficient and effective systems for managing such information; provides information services and support to Member States and to the IAEA; and assists with capacity building and training.

International Nuclear Information System (INIS)

INIS represents an extraordinary example of global cooperation, where 154 (IAEA, 2015) members allow access to their nuclear information resources, including grey literature, in order to preserve them for future generations and offer a freely available nuclear knowledge repository that can contribute to the world's sustainable development and further the use of nuclear energy for peaceful purposes. In addition to over one million full texts, more than 3.8 million bibliographic references to publications, documents, technical reports, non-copyrighted documentation, and other grey literature are made available. Overall, there are 800 GB of data in the INIS collection.
The role of INIS is to (a) Collect and process bibliographic metadata and full texts of nuclear literature; (b) Preserve NCL, such as documents, reports, theses, and other full text publications; (c) Disseminate publications from the INIS repository free of charge to all Internet users primarily through its website, which became free, open and unrestricted in April 2009 (IAEA, 2009).
Full text documents available from the INIS collection represent almost entirely nuclear related non-conventional literature (NCL) or grey literature. This PDF collection contains some very important historic and technical documents collected by INIS during the last 45 years (IAEA, 2015). Since much of this documentation originated in paper form, digitization of this collection was a huge project, converting millions of microfiche pages to electronic, fully searchable files. Optical Character Recognition (OCR) was performed on all documents within the collection, making it easy to index and search. Besides being a source of information when searching, the availability of full text gives INIS a special role in the area of nuclear information and

documentation — acting as the main custodian of this world information heritage and preserving this codified, specialized, scientific and technical knowledge.

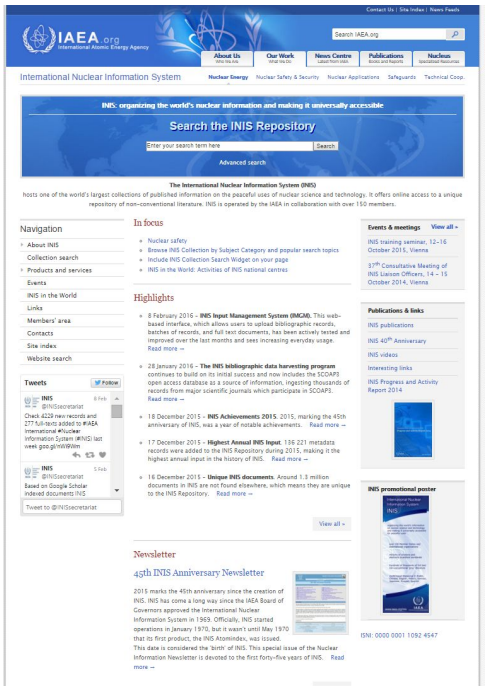


Figure 4: INIS website

INIS Repository

As of 1 November 2015, the INIS repository contains 3,860,194 bibliographic metadata records. Of those, 1,071,909 are full text documents, 751,198 of which are directly available from INIS, with the remaining 320,711 NCLs available from other sources. Only a small portion of the full text documents is restricted and kept for internal use.

On average, INIS adds around 120,000 bibliographic records and 13,000 full text PDF documents annually to its collection. The collection is accessible from the INIS website.

The INIS repository website[1] attracts many visitors and users. In 2015, there were over 1 million registered sessions and over 2 million pageviews, indicating 2 pages per session. Over 725,000 visitors came to the website with a bounce rate (single page visit) of 61%. At the same time, 30% of all visitors returned, most of them from Brazil, Canada, France, Germany, India, Japan, South Korea, the UK, and the US.

In 2014, there were 423,000 downloads, while in 2015, this number increased to 1.6 million, out of which 100,000 were registered through the INIS Collection Search (ICS)[2], with the remaining 1.5 million through Google.com or Google Scholar.

The most frequently used search terms in the ICS were nuclear reactors, radioactive waste, environment, Fukushima, and accidents. The most frequently downloaded documents through Google were those related to radiology and nuclear medicine, reactor safety, and radiation protection.

INIS collection by subject

The INIS collection covers around 50 well defined subject categories (IAEA, 2010) which are regularly maintained by INIS, and provides scope descriptions used by national and regional centres to categorize nuclear literature for INIS input. The INIS Joint Reference Series publications are also available on the INIS website.

The INIS collection covers all aspects of the peaceful uses of nuclear science and technology, such as nuclear reactors, reactor safety, nuclear fusion, applications of radiation and radioisotopes in medicine, agriculture, industry and pest control, as well as related fields of nuclear chemistry, nuclear physics and materials science. Special emphasis is placed on the environmental, economic and health effects of nuclear energy. Legal and social aspects associated with nuclear energy are also covered. Figure 5 lists a complete set of INIS Subject Categories.

[1] http://www.iaea.org/inis

[2] https://inis.iaea.org/search

S01 Coal, lignite, and peat
S02 Petroleum
S03 Natural gas
S04 Oil shales and tar sands
S07 Isotopes and radiation sources
S08 Hydrogen
S09 Biomass fuels
S10 Synthetic fuels
S11 Nuclear fuel cycle and fuel materials
S12 Management of radioactive wastes, and non-radioactive wastes from nuclear facilities
S13 Hydro energy
S14 Solar energy
S15 Geothermal energy
S16 Tidal and wave power
S17 Wind energy
S20 Fossil-fueled power plants
S21 Specific nuclear reactors and associated plants
S22 General studies of nuclear reactors
S24 Power transmission and distribution
S25 Energy storage
S29 Energy planning, policy and economy
S30 Direct energy conversion
S32 Energy conservation, consumption, and utilization
S33 Advanced propulsion systems
S36 Materials science
S37 Inorganic, organic, physical and analytical chemistry
S38 Radiation chemistry, radiochemistry and nuclear chemistry
S42 Engineering
S43 Particle accelerators
S46 Instrumentation related to nuclear science and technology
S47 Other instrumentation
S54 Environmental sciences
S58 Geosciences
S60 Applied life sciences
S61 Radiation protection and dosimetry
S62 Radiology and nuclear medicine
S63 Radiation, thermal, and other environmental pollutant effects on living organisms and biological materials
S70 Plasma physics and fusion technology
S71 Classical and quantum mechanics, general physics
S72 Physics of elementary particles and fields
S73 Nuclear physics and radiation physics
S74 Atomic and molecular physics
S75 Condensed matter physics, superconductivity and superfluidity
S77 Nanoscience and nanotechnology
S79 Astrophysics, cosmology and astronomy
S96 Knowledge management and preservation
S97 Mathematical methods and computing
S98 Nuclear disarmament, safeguards and physical protection
S99 General and miscellaneous

Figure 5: INIS subject categories

In order to review the INIS collection by subject area, 49 subject categories were grouped into 14 subject areas. The biggest number of documents in the INIS collection falls under the category of Economic, legal and social (8.2%). This is followed by Environment and earth science (5.7%), Elementary particle physics (5.1%), Engineering and instrumentation (4.7%), and Chemistry (3.4%). Figure 6 gives a more detailed breakdown of INIS collection subject categories and the number of documents contained in each one.

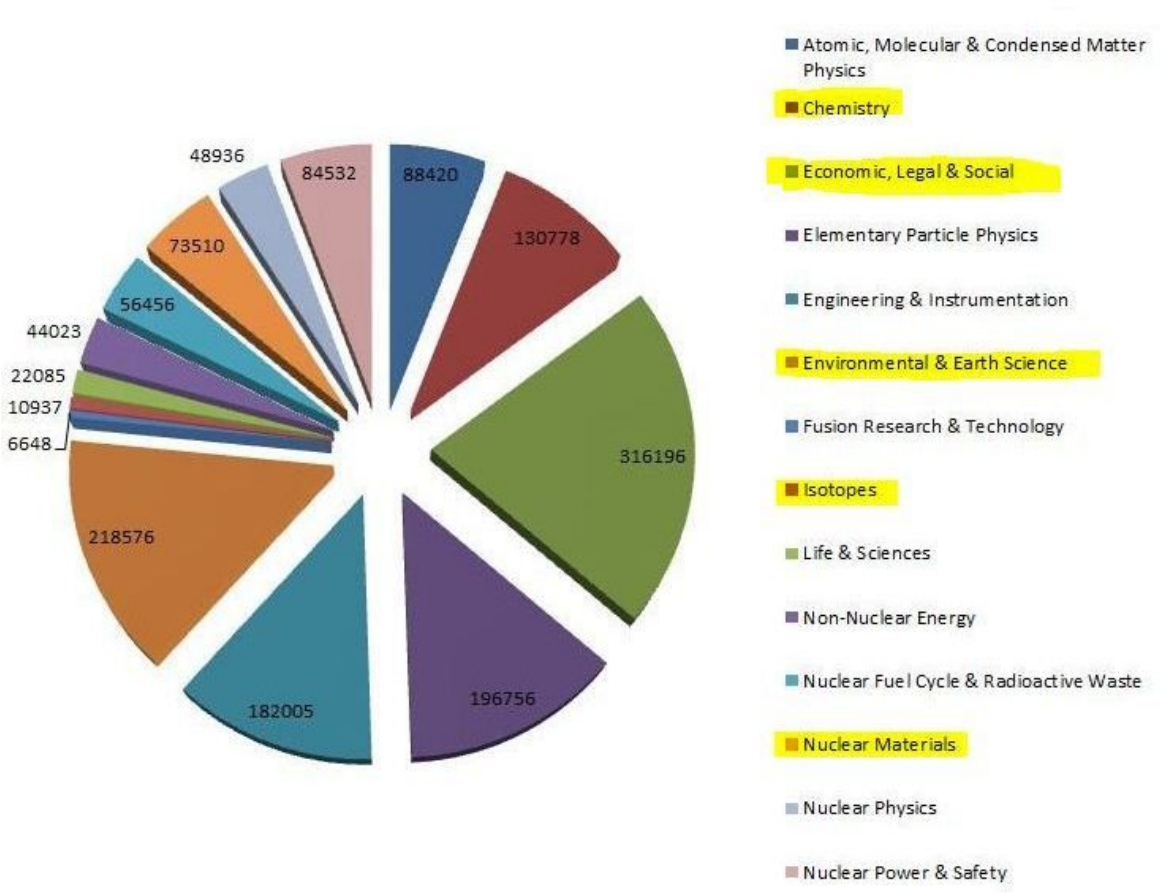


Figure 6: INIS collection by subject

INIS collection by record type

The INIS collection consists of seven types of literature – computer media, patents, books, reports, journal articles, miscellaneous and audio-visuals. The most populous are journal articles, comprising almost 62% of all records available in the INIS collection, followed by miscellaneous (11%) which include theses, pamphlets, brochures, conference proceedings, and reports (16%). Accordingly, it can be concluded that around 73% of the collection is represented by standard non-grey literature and only 27% is actually grey literature. However, it should be noted that while the INIS collection includes links to commercially available journal articles, a great majority of full texts are, in fact, grey, or non-conventional literature.

Figure 7 gives a more detailed break-down according to the various document types in the collection.

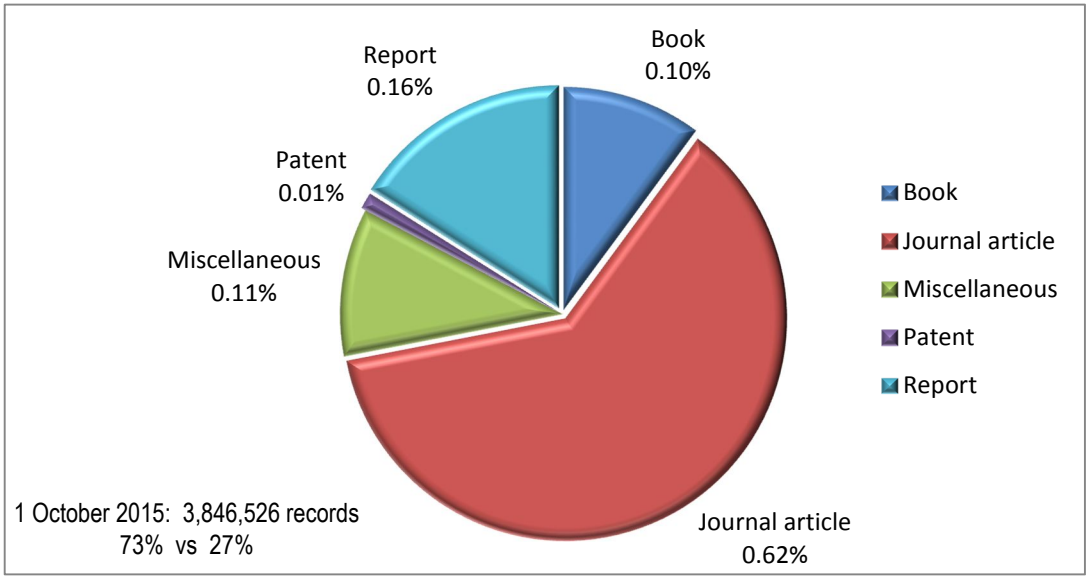


Figure 7: INIS collection by record type

Collection uniqueness

Of significance here is that the INIS collection is not indexed by commercial database providers such as Web of Science, EBSCO, Science Direct, ProQuest, etc. Although some benefits of having the collection included in one or any of these providers could be argued, a conscious decision was made to keep INIS outside of commercial channels and to continue offering it freely through the Internet and through WorldWideScience.org.

Analysis shows that around 80% of the INIS collection is uniquely available in Google.com solely from INIS. In order to verify this, a brief analysis was conducted using the most frequently used search terms – radiology, reactor safety, and radiation protection. The first 20 documents returned from each term in ICS using the search construction "exact phrase" + full text availability were then sought in Google.com using their exact full title, file type (PDF), excluding the iaea.org, worldwidescience.org and google.com websites. Radiology landed 2 outside links, while the other two had one link each. It was interesting to note that the links found were, in fact, the links from the originating publisher of the document, who also submitted the document to INIS.

Access analytics

As already established, journal articles represent the largest percent of the INIS collection (62%), followed by reports, miscellaneous, and patents. The last three types are grey literature and, combined, they make up 27.78% of the collection. Interestingly enough, the number of accesses, out of the total 1.3 million in 2015, when converted to a percentage, indicate almost the identical value (27.44%) as its representation in the collection, as shown in Figure 8.

	# of records (3,856,529)	% of the collection	# of records accessed	% of records accessed	# of accesses (1,296,607)	% of all accesses
Journal article	2,386,933	61.89%	108,074	4.53%	500,342	38.58%
Report	618,619	16.04%	60,260	9.74%	266,194	20.53%
Misc.*	406,386	10.54%	44,159	10.87%	211,631	16.33%
Book	390,067	10.11%	35,944	9.22%	304,192	23.46%
Patent	46,429	1.20%	2,212	4.76%	7,546	0.58%
Multimedia	7,829	0.20%	1,284	16.40%	6,498	0.50%
Obsolete types**	266	0.00%	19	7.14%	104	0.00%

27.78%

27.44%

Figure 8: INIS collection access statistics by document type

* Miscellaneous: theses, pamphlets, brochures, conference proceedings
** Obsolete: G: Maps; F: Audio-visual materials; C: Conference

It is interesting to note that, although journal articles make up the biggest part of the collection, only 4.5% of them were accessed compared to multimedia (16.4%), miscellaneous (11%) and reports (almost 10%).

Conclusions

Returning to the starting assumption that the reasons for public interest lay in three main features of the collection, (subject coverage, type of documents and uniqueness), and based on the above analysis, a number of interesting conclusions can be drawn.

Open access collections, repositories and databases can attract a huge number of users, especially if the subject is of broad interest. However, even specific subject related coverage collections attract special user groups in bigger numbers. In the case of INIS, these are the Internet domains .edu and .gov.; the education sector because of interest in research and development and the government sector mainly because of relevant administrative documentation and policy documents.

It is also interesting to note that the most frequently downloaded documents are NOT from the most populous parts of the collection. Although journal articles make up the greatest portion of the INIS collection, grey literature (reports and miscellaneous) is the most frequently accessed.

Discovering that around 80% of the INIS collection is unique was a large incentive to continue collecting the same type of information and documentation and to further enhance INIS as an international collaborative effort. Related to that were the findings that when full text NCL is available on websites other than INIS, it is usually only on the originator's website.

It can also be concluded that the percentage of accesses to NCL (part of the total number of accesses) is directly proportional to their representation in the collection. However, as already stated, the total percentage of records accessed is much higher, representing special interest in that part of the collection.

Finally, it seems that users are looking for information irrelevant of its form and that they gravitate mainly towards full texts, which might explain why multimedia is so frequently downloaded and why there is twice as much interest in accessing reports as in journal articles.

References

IAEA. (1956). *The Statute of the IAEA*. Získáno 2015, z https://www.iaea.org/about/statute

IAEA. (3. April 2009). *Opening of the access to the INIS Database on the Internet*. Získáno 2015, z https://www.iaea.org/inis/highlights/2009/news_2009_03.html

IAEA. (2010). *Subject Categories and Scope Descriptions*. Získáno 2015, z http://nkp.iaea.org/INISSubjectCategories/

IAEA. (2015). *IAEA Annual Report for 2014*. Získáno 2015, z https://www.iaea.org/Publications/Reports/Anrep2012/orgchart.pdf

IAEA. (2015). *INIS Membership*. Získáno 2015, z https://www.iaea.org/inis/about-us/membership.html

IAEA. (9. November 2015). *Nuclear Information Newsletter - 45 years of INIS*. Načteno z https://www.iaea.org/inis/highlights/2015/news-20151109.html

IAEA. (2015). *The "Atoms for Peace" Agency*. Získáno 2015, z https://www.iaea.org/about/about-iaea

Extracting Value from Grey Literature: processes and technologies for aggregating and analyzing the hidden "big data" treasure of organizations

Gabriele Motta, Roberto Puccinelli, and Lisa Reggiani, Network and Information System Office, CNR;
Massimiliano Saccone, Central Library, CNR, Italy

Abstract
Grey literature can be a valuable source of information about organizations' activities and, to a certain extent, about their identity. Some of the major problems that hinder their full exploitation are the heterogeneity of formats, the lack of structure, the unpredictability of their content, the size of the document bases, which can quickly become huge.
The collection and mining of grey literature can be applied to individual organizations or classes of organizations, thus enabling the analysis of the trends in particular fields. To this end, some techniques can be inherited from the best practices for the management of structured documents belonging to well identified categories, but something more is needed in our case.
Obvious steps are: identifying sources, collecting items, cleansing and de-duplicating contents, assigning unique and persistent identifiers, adding metadata and augmenting the information using other sources. These phases are common to all digital libraries but further steps are required, in our opinion, in the case of grey literature in order to build document bases of value. In particular, we think that an iterative approach would be the most suitable in this context, one including an assessment of what has been collected in order to identify possible gaps and start over with the collection phase.
We think that big data technologies, together with information retrieval and data and text mining techniques, will play a key role in this sector. This "bag of tools" will certainly facilitate the management, browsing and exploitation of large document bases that belong not only to a single organisation but also, for example, to a large number of organizations working in a particular sector. This on the one hand opens new scenarios regarding the type of information that can be extracted, but on the other hand introduces new problems regarding the homogenization of contents, formats and metadata, and additional issues related to quality control and confidentiality protection. We believe that in this context the incremental-iterative approach would help address, gradually and based on real cases, the problems mentioned above.
In this paper we describe the process that, in our opinion, should be put in place and a high level ICT architecture for its dematerialization, along with the technologies that could be leveraged for its implementation.

1. Introduction

Heterogeneity, complexity and structural variability characterise the 'Big data' concept and related contexts, as well as the grey literature system and its environment. It seems then natural to leverage 'Big data' technologies not only to manage grey literature contents and metadata, but also to analyse them in order to extract useful and non-trivial information, which may be hidden in them. Current 'Big data' technologies, in fact, satisfy base requirements such as scalability, flexibility (e.g. ability to cope with different formats), interoperability, openness.
On the methodological side, it seems to us appropriate to use a practical approach (i.e. non theoretical), based on some key elements:

- sharing of a fundamental language;
- collaboration between actors and communities involved;
- interoperability framework.

In the following we present our view about the analysis of grey literature contents and metadata by means of 'Big data' solutions. We start by discussing the issues related to providing a formal definition of 'Big data' and then present a short survey of software platforms and tools. We then highlight the connections between 'Big data' and grey literature and illustrate a methodological approach for combining the two concepts. We finally describe our proposal of an integrated system for the management and analysis of grey literature contents and metadata.

2. Big data: definition and related issues

2.1. Original heterogeneity and massive diffusion

The term "big data" has emerged from activities in different fields and has over time become more and more ambiguous mainly due to:

- *heterogeneity as original pattern* - the term derived from different contexts (research, industry and communication media in general) [1];
- *pervasive diffusion in many different usage contexts* - it has massively spread in very different contexts (society and public sector, business & research); examples of big data may indeed include social media data, e-business data, linked data, web usage data, government open data, research data, as well as data created in such diverse environments as cloud-based computing infrastructures, virtual collaborations and projects, e-science, e-humanities and e-social sciences.

Big data have become a cultural shift in data processing and also the preferred solution for extracting the maximum benefit from data made available by extremely heterogeneous sources. The related technologies seem to cope effectively with the information overload typical of our era [2].
Moreover, the growing usage of insights generated through Big Data Analytics in companies, universities and institutions has modified decision-making culture as well as scientific method [3].
Pervasiveness and intrinsic vague polysemy have produced a **strong semantic interference**: terms more or less synonymous for 'big data' have contaminated it at the connotative level and sometimes at the denotative level, enriching it with additional meanings or favouring its use in hybrid and different ways, so that they have been legitimated coming into common use. There are indeed many different terms used in literature that may refer to or be associated with the phenomenon of 'big data', including such terms as *digital data, research data, linked data, open data, web of data and data repositories* [2].
The multiple meanings attributed to the 'big data' concept are connected to the different production and use contexts, to the involved actors and their purposes. These definitions also show a substantial and significant diachronic dimension, changing over time in accordance with the cultural and technological evolution that characterizes application fields and user communities.
Attempts to provide operational definitions – dependent on actors, goals and circumstances and gradually refining themselves – have proliferated. They are characterized by a certain "methodical negation": approaches, methods, tools and technologies used until now would be, according to those definitions, inadequate to exploit such data whose volume is too big, whose production rate is too high and which are too heterogeneous and complex [4].

2.2. Big is a moving target[1]: the 3 "V"s

The three characteristics that recur in all definitions are:

- *volume (scalability)*: the size in this case does not refer to absolute values, but is related to contexts, tools and to intrinsic constraints imposed by human-computer interaction;
- *velocity (timeliness)*: i.e. the speed at which data is generated, retrieved, and need to be processed; it is a relative concept constantly evolving, too;
- *variety*: i.e. heterogeneity - of sources, types, internal structures, formats, etc. - and complexity, linked to different contexts, instruments, procedures and purposes of production, collection, management and reuse of data.

Most of the frequently mentioned definitions include and/or refer to these three characteristics, starting from a MetaGroup (now Gartner) report (2001). The Gartner report made no mention of the term "big data" and exclusively used the general term "data" with reference to data management and data management challenges, emphasizing the relationship with e-commerce. It repeatedly alluded to huge and continuously growing amounts of data and highlighted the increasing size of data, the increasing rate at which they are produced and the increasing range of adopted formats. Gartner proposed a threefold definition encompassing the "three Vs":

[1] Nature Special Issue, 4th September 2008 - http://www.nature.com/news/specials/bigdata/index.html (focusing on big data and their implications for science) [5].

Volume, Velocity, Variety, predating the current trend [6]. The three features, crystallizing over time, are always associated – in whole or in part – with the 'big data' notion.
Given the high variability and volatility of the data produced in many fields of interest, a fourth feature has progressively gained a significant relevance: veracity, a characteristic required to provide a solid ground for information extraction and effective data reuse. It is tightly tied to quality of data, data sources and processes – concepts that can be summed up by the terms 'accuracy', 'reliability', 'traceability' and 'trustability'. A recent survey on big data (2015) provided the following definition:
"Veracity (Data in doubt). Veracity is what is conform with truth or fact, or in short, Accuracy, Certainty, Precision. Uncertainty can be caused by inconsistencies, model approximations, ambiguities, deception, fraud, duplication, incompleteness, spam and latency. Due to veracity, results derived from Big data cannot be proven; but they can be assigned a probability [7]."
Therefore, within the four characteristics above described preferential relationships can be highlighted between:

- volume (scalability) & velocity (timeliness): *quantitative*, i.e. measurable aspects, apparently objective and thus more easily definable and manageable;
- variety & veracity/validity: *qualitative*, i.e. not measurable aspects, difficult to control, to define and to use [4].

The stress on the first two properties derives, at least partially, from the original inception and use of big data within business and commerce, hard sciences and ICTs sectors. This stress has not only persisted but has also increased over time: technologies developed for data analysis as well as control of data useful to identify choices, habits and behaviours of consumers/users have predictably focused on big data quantitative measurable dimensions, that continue to play a critical role in the current business models.
To sum up, the 'big data' concept consists of: data, relation contexts, technologies and tools that ensure their use.

3. Big data and Technologies

This section presents a brief survey of the most widely used Big Data technologies. For further details, see [16] and [17].

3.1. Big Data storage and processing technologies

As regards "Big Data" storage and processing, the first word that comes to mind to insiders is "Hadoop". This is an Apache project mainly focused on developing four modules:

- *Hadoop Common*: the common utilities that support the other Hadoop modules.
- *Hadoop Distributed File System (HDFS™)*: a distributed file system that provides high-throughput access to application data.
- *Hadoop YARN*: a framework for job scheduling and cluster resource management.
- *Hadoop MapReduce*: a YARN-based system for parallel processing of large data sets.

The framework is written in Java and its goal is to provide the ability to store and process large amounts of data using commodity hardware. One of its most notable features is to provide high availability by detecting and handling failures at the application layer, thus freeing from the need for complex hardware-based solutions.
For the purposes of this paper, the two modules of interest are HDFS and MapReduce, which are described in detail below.

HDFS. It is a distributed file system designed in reference to Google File System (GFS) in order to run on large clusters of commodity hardware. As described in [16] "HDFS is the basis for main data storage of Hadoop applications, which distributes files in data blocks of 64MB and stores such data blocks in different nodes of a cluster, so as to enable parallel computing for MapReduce. An HDFS cluster includes a single NameNode for managing the metadata of the file system and DataNodes for storing actual data. A file is divided into one or multiple blocks and such blocks are stored in DataNodes. Copies of blocks are distributed to different DataNodes to prevent data loss." The HDFS interface has been designed based on a similar approach to that of the Unix file system.
MapReduce. This framework implements the homonymous programming model, which is currently one of the most widely used for processing large datasets. It includes the ResourceManger module, running on the master node of the cluster and orchestrating the computational resources, the NodeManager module, running on the slave nodes (one instance

per node) that perform the actual work, and the MRappMaster module (one per application). A MapReduce job handles input and output data sets in the form of <key, value> pairs, typically stored in a distributed file-system. The MapReduce framework and the HDFS generally run on the same set of nodes, thus enabling to dispatch jobs to the nodes where data reside. This approach reduces the bandwidth usage.
In the MapReduce paradigm, the Map function provides intermediate outputs as sets of <key, value> pairs and the Reduce function merges the intermediate values with the same key.
As stated in the official MapReduce tutorial "… Minimally, Hadoop applications specify the input/output locations and supply map and reduce functions via implementations of appropriate interfaces and/or abstract-classes. These, and other job parameters, comprise the job configuration. The Hadoop job client then submits the job (jar/executable etc.) and configuration to the ResourceManager which then assumes the responsibility of distributing the software/configuration to the slaves, scheduling tasks and monitoring them, providing status and diagnostic information to the job-client."[2].

It is worth mentioning that MapReduce is not the only Big Data processing paradigm. Other programming models are [16]:

- *Dryad*, whose operational structure is represented as a directed acyclic graph, in which vertexes represent programs and edges represent data channels;
- *All pairs*, that focuses on comparing element pairs in two datasets by a given function;
- *Pregel*, where each computational task is expressed by a directed graph constituted by vertexes and directed edges, in which every vertex is related to a modifiable and user-defined value and every edge is related to its source vertex and is constituted by a modifiable and user-defined value and an identifier of a target vertex.

3.2. NoSQL Data bases

The term "NoSQL" generically refers to database technologies not based on the relational model. It is by some sources presented as an acronym standing for "Not only SQL", thus implying that, in this context, the use of the SQL is neither excluded nor mandatory.
This set of technologies has been developed to overcome the limits of the relational model, mainly in terms of horizontal scalability and/or performance. There are several types of NoSQL databases, each one focused on or more suited for a particular set of problems. A particular NoSQL database may be faster than an RDBMS for some operations and slower for other operations.
We do not present in the following a complete survey of NoSQL databases but just provide a list of the main types, each one with a short list of examples.

3.3. Key value databases

In this type of data base, data are stored in the form of <key, value> pairs. The underlying data model is the associative array, i.e. a set of <key, value> pairs where keys are unique. Pairs can be ordered lexicographically to improve performance of interval queries and values can be retrieved based on the corresponding keys. Examples of KV stores are :
Amazon Dynamo, a highly available and expandable distributed key-value data storage with a simple read-write interface. Flexibility and high-availability are achieved through data partitioning and replication. Consistency across replicas is guaranteed as an asymptotic condition (eventual consistency);
Voldemort, developed for LinkedIn, is a KV store where key words and values can be composite objects constituted by tables and images. Its interface includes three simple operations: reading, writing, and deletion. Voldemort does not ensure data consistency but supports optimistic locking for consistent multi-record updating. Data can be managed in RAM or via storage engine (Berkeley DB and Random Access Files).
Other examples of KV stores are Redis, Tokyo Canbinet and Tokyo Tyrant, Memcached and MemcacheDB, Riak and Scalaris.

3.4. Column-oriented databases

As suggested by the name, databases of this type store and process data according to columns other than rows. Columns and rows are segmented in multiple nodes to realize expandability

[2] https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html

[16]. In a relational database system, tables are composed by rows, each one usually identified by a primary key. Example:

Employee table:

RowID	First Name	Middle Name	Last Name	Department
0001	Michael	John	Stanford	Sales
0002	Mary	Ann	Naughton	R&D
0003	Michael	Bradley	Rutherford	Operations
0004	David	Robert	Jones	R&D

In a columnar database the above table could be stored as:

Michael: 0001, 3; Mary: 0002; David: 0004;
John: 0001; Ann: 0002; Bradley: 0003; Robert: 0004;
Stanford: 0001; Naughton: 0002; Rutherford: 0003; Jones: 0004;
Sales: 0001; R&D: 0002, 0004; Operations:0003;

Using this model, queries like "find all employees with First name 'Michael'" would be performed with a single access. In addition, aggregated queries, like the ones usually performed in data warehouses, can benefit from this type of layout.

3.5. Document Databases

Databases of this type store records as documents encoded according to several standards (JSON, BSON, XML, YAML, etc.). Each document is identified by a unique key. They generally allow for data distribution and redundancy over clusters of servers in order to achieve horizontal scalability, high availability and fault tolerance.
Examples of document database are MongoDB, SimpleDB and CouchDB.
MongoDB is an open source platform where documents and queries are both represented as BSON and query syntax is similar to JSON. Automatic sharding distributes data across a cluster of machines. Eventual consistency is guaranteed and Multi-Version-Concurrency-Control can be implemented on top of the MongoDB features.
SimpleDB is an Amazon Service accessible via web service API. It is a distributed database but does not support automatic partition. It allows users to use SQL to run query. Eventual consistency is guaranteed but not MVCC.
Apache CouchDB is an open source platform where data are stored as JSON. It features a RESTful HTTP API for querying and updating documents. Each document is uniquely identified. "If a document needs to be modified, the client can download the entire document, modify it, and then send it back to the database. After a document is rewritten once, the identifier will be modified and updated". [vd fonte] CouchDB features eventual consistency and implements synchronous MVCC. ACID The CouchDB file layout and commitment system also features all Atomic Consistent Isolated Durable (ACID) properties. It also features a view model that allow aggregation and reporting on the documents in the database. Views are defined via Javascript as a Map function in a MapReduce system.

3.6. Triple stores

A triple store keeps information in the form of (subject, predicate, object) triples and is generally suited for semantic applications. Triples can be represented in RDF (Resource Description Framework) format and data can be retrieved via a query language (e.g. SPARQL). Examples of triple stores are AllegroGraph, Apache Jena, Oracle NoSQl Database, SparkleDB, Virtuoso Universal Server.

3.7. Other types of NoSQL databases

Other types of NoSQL databases are Graph databases, Object databases, Tabular databases, Tuple stores.

3.8. Data warehousing

When it comes to heterogeneous data sources, one possible approach to the collection and analysis is to apply adequate procedures that extract data from the original repositories, process them according to the needs, and finally, load them into appropriate structures, optimized for aggregated queries (ETL procedures). Data warehousing tools in the Big Data field are:

- *Hive*, a software that allows to project a unified structure onto distributed heterogeneous data and to query them using an SQL like language;
- *Pig*, a platform that allows to compile and run MapReduce programs written in a textual

language called Pig Latin;
- *Sqoop*, a tool designed for efficiently transferring data between structured, semi-structured and unstructured data sources.

3.9. Big Data Analytics tools

In 2012 KDNugget conducted a survey among 798 professional asking the following question:
"What Analytics, Data mining, Big Data software you used in the past 12 months for a real project?"
The top five answers were [16]:
- *R*, an open source execution environment for programs written in the homonymous language;
- *Excel*, the only commercial software among the top five;
- *Rapid-I Rapidminer*, a tool that allow to define and execute complex analysis jobs described as xml files and visualized through an intuitive GUI;
- *Knime*, written in Java and based on the Eclipse framework, features a visual GUI and can be easily extended via plugins;
- *Weka*, written in Java and integrated with the Pentaho BI platform, provides such functions as data processing, feature selection, classification, regression, clustering, association rule, and visualization.

For the purposes of this paper, also tools for text mining that can perform document clustering and classification are of interest. Examples are:
- *Rapidminer* with its Text Processing Extension;
- *Knime* Text Processing extension;
- *Carrot2*.

4. Big data and Grey Literature

With regards to the grey literature big data include [8]:
- research data (in their processing stages, starting from the raw data set acquisition), textual and not-textual documents and virtual data representations;
- the contexts and their relationships, typical of the complexity of contemporary science, of its various actors and communities;
- infrastructures, instruments, tools and ICT methods.

Variety – as heterogeneity and complexity – and, more generally, structural variability that characterize big data (i.e. data, context & relationships, instruments & tools) emphasize the intrinsic critical features of grey literature.

4.1. Benefits and hidden value: the synergy between big data and grey literature

The *quality* dimension in its articulation – above highlighted for big data – has become decisive in grey literature and in its connection with big data, because on the one hand it enables to increase and fully exploit grey literature potential and on the other it allows to effectively cope with actual shortcomings and risks. In fact, it drives the various stakeholders and communities to improve all the steps of process, (acquisition, processing, dissemination) and to focus in particular on data and metadata quality control [9].
Taking into account the big data characteristics and adequately valuing their contexts and relationships, it is possible to:
- improve communication between actors and communities;
- discover more easily implicit research trends and/or latent themes, both at the individual and community level, especially inter-, trans- and multi-disciplinary ones[10];
- foster multi-, trans- and inter-disciplinary research;
- create and strengthen effective antidotes – approaches, methods and practices – to hyper-specialization, that always means 'non-communication' and mutual incomprehension [11].

The discovery of hidden research trends and latent topics [11] can be useful to various actors, in particular:
- to researchers, to identify mainstream research trends, before they emerge to the surface, and to detect upcoming/implicit research trends or interesting connections between research groups/fields;

- to policy makers, because it supports them in long term research investment planning[3];
- to industries, because it anticipates innovation, fostering it; thereby it accelerates and facilitates convergence of research and innovation, a key trend in the current national and European planning.

5. Methodological approach

From a methodological perspective, in order to take the most advantage from grey literature and big data interaction, it is necessary to focus on some interconnected aspects of quality:

- transparency in methodology, experimental observation, and data collection;
- reliability and reusability of scientific data and research output;
- accessibility and transparency of scientific communication processes;
- adoption of web-based tools to facilitate scientific collaboration and to enhance process accessibility and transparency.

5.1. Context and reference framework

In order to achieve the above listed goals it is necessary to define:

- the conceptual framework and methodology;
- the fundamental processes involved, with particular attention to those of scholarly communication;
- a common language among stakeholders and research communities that can go beyond peculiarities of individual entities (institutions/companies/persons) and groups (disciplinary and/or national communities, teams, research networks and communities…).

The *Reference Framework* we propose for Big Data and scholarly communication (which includes Grey Literature) is complex and heterogeneous, and presents different aspects:

- multidimensionality: it regards different sectors and settings of intervention, which are developed at different levels;
- interdisciplinarity and multidisciplinarity;
- diachronicity: it characterises the influential factors and their mutual interdependencies.

5.2. A methodological road map

The above listed aspects should, in our opinion, be addressed by:

- identifying the different stakeholders (researchers, policy makers, ICT and information specialists, citizen, industries, libraries, etc.), and their related purposes and strategies;
- identifying and knowing the different information and cultural contexts, practices, approaches, channels and specific jargons (technical jargon);
- identifying and using consistently the instruments/tools that allow to plan and manage the actions in the R&D system at multiple levels and in different intervention areas (International, European and National policies and programmes, guidelines; laws & rules; institutional regulations; agreements, contracts, etc.);
- verifying implementation of the interventions and their effectiveness at different levels, through monitoring and evaluation systems;
- as a consequence, guiding the choices on strategies, instruments, actions at different levels and environments on the basis of the analysis previously carried out (by confronting and relating results).

5.3. Interoperability

In order to act on processes, it is necessary to define a common interoperability framework, i.e. to define a common multi-level and multi-dimensional framework in which main stakeholders and different research communities can share objectives, resources, infrastructures, instruments/tools, services and results.

The strategic intervention areas are the following: political, institutional and economic setting; legislative, legal and regulatory framework; organizational setting; technical and technological environment.

However, firstly, culture, meaning and language sharing is the basis of all the interoperability strategies. In fact, this approach preliminarily requires the use of a common language among the different contexts and actors, that is, a common cultural code.

[3] Big data usage requires that stakeholders are aware of all the legal and ethics-related issues (privacy, security, intellectual property and copyright, etc.) [12].

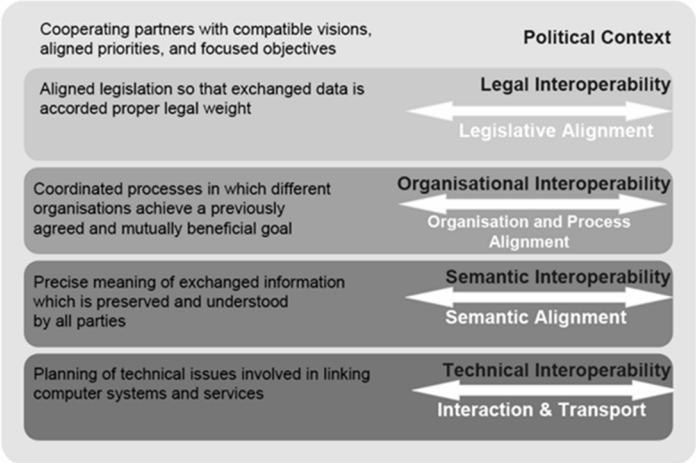


Figure 1 - European Interoperability Framework (EIF) [13]

5.4. A first proposal

The different issues, quickly examined in the previous section, are addressed by a first methodological proposal, which includes three sequential steps:

1. to address key challenges, regarding the entire big data life cycle and their production, collection, management and re-use systems, through a multi-dimensional analysis, that methodically involves different communities and actors;
2. to build a stable collaboration between actors and communities, characterized by very different cultures and interests, often competing with each other (i.e. to understand complexity and diversity in order to maximize the heterogeneity potential);
3. to exploit Linked Open Data as context of relationships, in order to promote the exchange of good practices and to improve the transparency of processes.

6. Proposed technical solutions

In the following sections, we illustrate our proposal of an integrated system for the collection, processing, analysis and dissemination of grey literature contents coming from different communities. Such system, in our opinion, allows extracting useful information from the collected data and metadata, thus allowing the full exploitation of their potential. We start from the analysis types we deem interesting for our purposes and then present a high level logical view of the architecture. Finally, we illustrate the workflow that should be implemented in order to acquire, cleanse, enrich and process the items of interest.

6.1. Analysis types

There are many types of analysis that can be performed on a document base of the kind addressed in this paper. In order to focus our activities on the needs of the stakeholders listed in the previous paragraphs (political decision makers, industry, researchers), we decided to initially restrict our attention to a limited number of analysis, whose potential benefits are more obvious. In first place, we think that the heterogeneity of document types, formats and sources poses serious problems of classification and in general of metadata completeness. To this end, we think that applying well-selected classification techniques might be of help. Machine learning algorithms of this type take as input:

- A document base;
- A list of categories;
- A training set (i.e. a set of documents for which the classification has already been performed and is assumed to be correct);

and assign each document to one category. This process may be used to fill some metadata fields whose values are missing.

Another interesting use of machine learning algorithms could be the application of clustering techniques for the detection of non-trivial connections between research fields or organizations/research groups. Let us assume to have partitioned a document base into clusters based on the textual content. We could then evaluate the relative frequencies of the values of particular metadata fields within each cluster in order to identify patterns and/or correlations.

As a last example, we would like to propose the application of association rules to detect interesting patterns. When applied to the grey literature, these may be useful for the early

detection of underlying trends that may not (yet) be evident from journal articles or conference proceedings papers. As a result of such analysis we may, for instance, discover that, if a research group is specialized in field A and is based in the geographic area B, then they have a strong probability to participate to initiatives in the field C. For further details about the above-described algorithms, see also [14] and [15].
As a final remark, we would like to point out that analyses may be performed both on metadata and contents and that it is also important in a system as the one we are proposing to have classical Business Intelligence tools in order to perform normal aggregate and detail reporting.

6.2. Proposed architecture

In this section, we present a high level logical view of a system for the acquisition and analysis of grey literature contents. This system is designed to serve different user communities and to acquire contents in different formats from providers who have signed agreements that bind them to respect some basic requirements in terms of content quality. It is a multi-layer architecture featuring the "Data layer" at its base. This is the level where contents, metadata and analysis results are kept, each one in the appropriate type of storage. We think that contents should be stored in a distributed file system whereas metadata are better managed using a NoSQL database, (e.g. an eventually consistent Key-Value Store like Cassandra). Analysis results may be stored either in a NoSQL database or in a relational one. A triple store is included for semantic data.
On top of the "Data layer", there is the "Access layer", which provides access to metadata and contents in both input (Ingestion) and output (Retrieval). As regards ingestion, a high throughput, multi-threaded, queue-based collection system is required, as the one developed for the "Science and Technology Digital Library" project [18], successfully completed by AgID and CNR. Such a component feeds a cleansing and enrichment workflow that prepares metadata and contents for the analysis. As regards retrieval, search functions are implemented at this level using components like Apache Solr. Search and browsing functions are exposed to the upper layer as RESTful API.
The "Service layer" includes components for text and data mining as the ones presented in the previous sections. In addition, classic BI tools are included for normal reporting. At the top of the stack is the Presentation layer, which includes portals for the access of contents, analysis results and reports.
As we said before, different user communities may access this system, thus requiring access control mechanisms and user profiling functions.

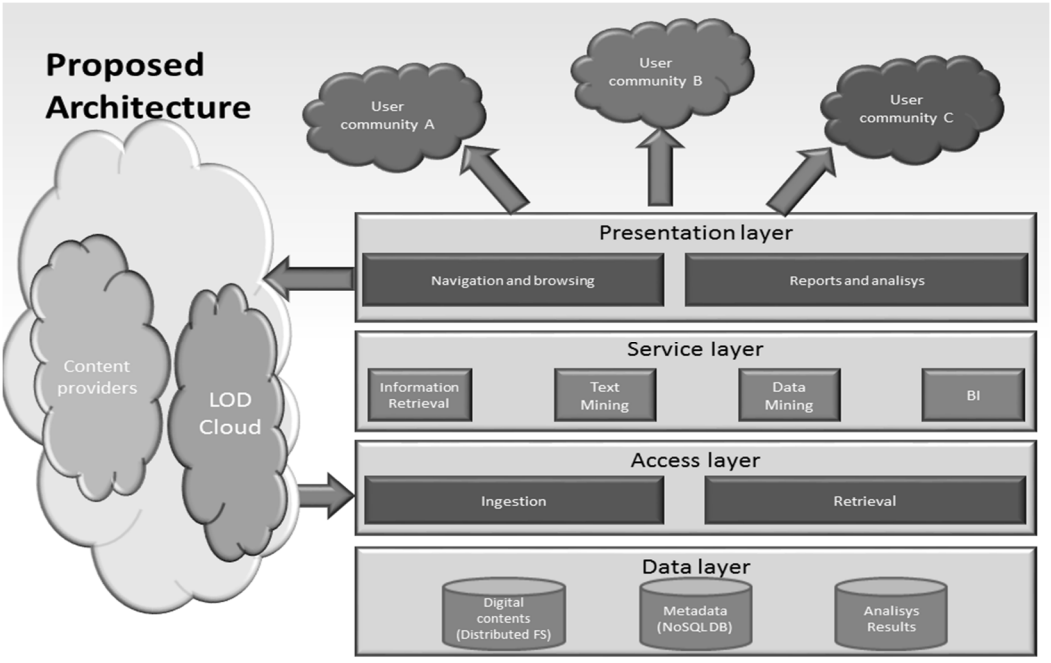


Figure 2 - Proposed architecture (logical view)

6.3. Proposed workflows

As regards the ingestion and processing workflow, contents are submitted to the system as Submission Information Packages (SIP). For each SIP, contents and metadata are extracted. Metadata are cleansed and enriched using inference algorithms, external sources and content-bundled metadata (i.e. metadata enclosed in the content files as in PDF, Word, JPEG, etc.). Finally, metadata are fed to the data mining algorithms.

On the other hand, contents pass a virus check and, in case of textual documents, a text extraction step. A deduplication algorithm helps detecting multiple copies of the same item. Contents too are fed to the data and text mining algorithms. Search indexes are periodically refreshed to ensure content retrieval.

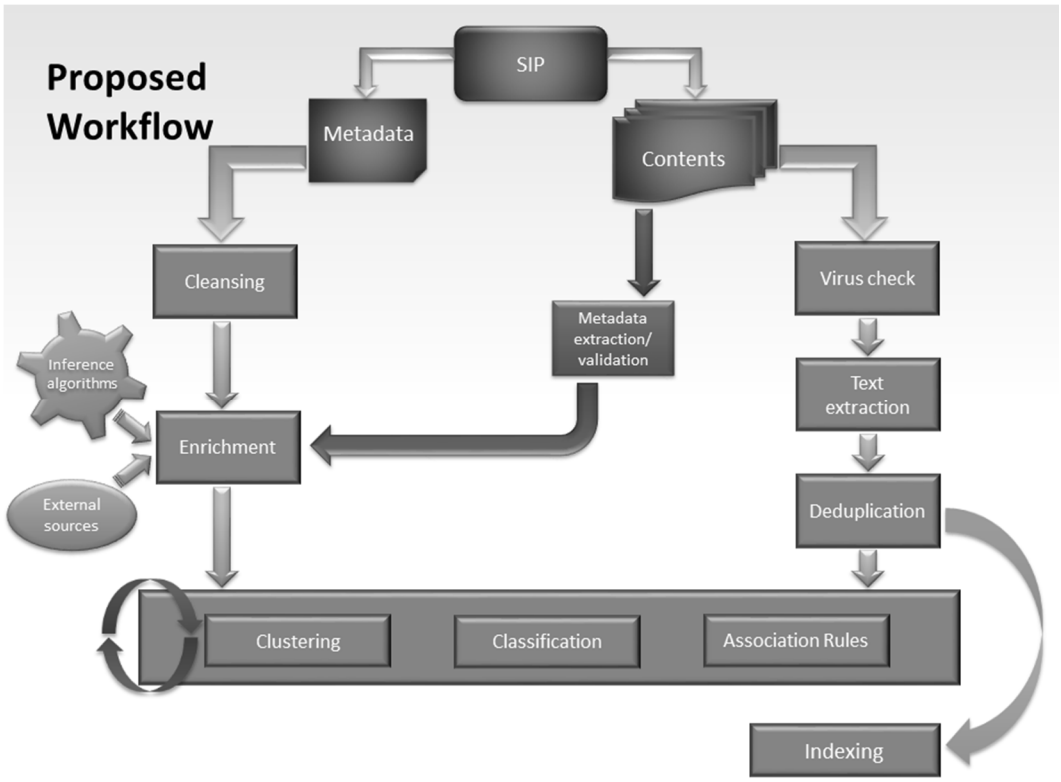


Figure 3 - Proposed core workflow

7. Conclusions and future work

In this paper we have presented a methodological framework for the management, processing and analysis of grey literature contents and metadata, along with a high-level ICT architecture based on 'Big data' technologies. The methodological framework can, in our opinion, be useful to tackle the main issues regarding the addressed problems, mostly related to organisational and communication aspects, and in particular to cope with interoperability concerns.

It can also contribute to effectively promote the network of relationships among the main actors and communities.

This approach, in our view, favours a) the integration of metadata, contents and analysis results in the LOD cloud and b) the exploitation of feedback from user communities for the continuous improvement of analysis result quality.

Moreover, despite the inherent differences of cultural approaches, it makes possible the gradual convergence of the involved actors about some key definitions and related evaluation criteria.

On the technical side, one important issue that will be addressed in our future activities regards the criteria for the selection of contents and metadata to be analysed. It is indeed a non-trivial task in such a heterogeneous context to ensure at the same time richness and meaningfulness of the information content and absence of noise.

References

[1] Jonathan Stuart Ward, Adam Barker, *Undefined By Data: A Survey of Big Data Definitions*, eprint arXiv:1309.5821, September 2013

[2] Ali Shiri, *Linked Data Meets Big Data: A Knowledge Organization Systems Perspective*, Advances In Classification Research Online, vol. 24, issue 1 (2014), p. 16-20

[3] Andrea De Mauro, Marco Greco, Michele Grimaldi, *What is big data? A consensual definition and a review of key research topics*, AIP Conference Proceedings, vol. 1644, issue 1 (2015), p. 97-104

[4] H.V. Jagadish, *Big Data and Science: Myths and Reality*, Big Data Research, vol. 2, issue 2 (2015), p. 49-52

[5] Nature Special Issue, 4th September 2008 - http://www.nature.com/news/specials/bigdata/index.html

[6] Douglas Laney, *3D data management: Controlling data volume, velocity and variety*, META Group, February 2001

[7] Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle, *Understandable Big Data: A survey*, Computer Science Review, vol. 17 (2015), p. 70-81

[8] Julia Gelfand, Daniel Tsang, *Data: Is it Grey, Maligned or Malignant?*, International Journal on Grey Literature, vol. 11, issue 1 (2015), p. 29-40

[9] Luisa De Biagi, Roberto Puccinelli, Massimiliano Saccone, Luciana Trufelli, *Research product repositories: Strategies for data and metadata quality control*, GL Conference Series, vol. 13 (2012), p. 132-142

[10] Arun S. Maiya, Robert M. Rolfe, *Topic similarity networks: Visual analytics for large document sets*, in 2nd IEEE International Conference on Big Data, IEEE Big Data 27-30 October 2014; Washington; United States, p. 364-372, IEEE, 2015

[11] Gali Halevi, Henk F. Moed, *The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature*, Research Trends, issue 30 (September 2012), p. 3-6

[12] John Carlo Bertot, Heeyon Choi, *Big Data: Strategies, Issues, & Recommendations White Paper*, Information Policy & Access Center, 2012

[13] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions *Towards interoperability for European public services*, European Commission, 2010

[14] Max Bramer, *Principles of Data Mining*, Springer-Verlag London Limited, 2007

[15] Pieter Adriaans, Dolf Zantinge, *Data Mining*, Addison Wesley Longman Limited, 1998

[16] Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung, *Big Data Related Technologies, Challenges and Future Prospects*, Springer, 2014

[17] Alessandro Rezzani, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Maggioli Editore - Apogeo Education, 2013

[18] Edgardo Ambrosi, Gabriele Motta, Roberto Puccinelli, Alberto Salvati, Marco Spasiano, *Science and Technology Digital Library Software Architecture Document*. Science & Technology Digital Library project deliverable.

Leveraging Grey Literature – Capitalizing on Value and the Return on Investment: A Cumulative Case Study

Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands
Stefania Biagioni and Carlo Carlesi, ISTI-CNR, National Research Council, Italy
Christiane Stock, Inist-CNRS, National Centre of Scientific Research, France

Introduction

Leveraging grey literature not only seeks to capitalize on its value for science and the public good, but also anticipates a return on investments in some measure. In an attempt to assess GreyNet's return on investment in grey literature, this study identifies and reviews efforts made to leverage its own information products and services, since its relaunch in 2003.

The method of approach is a cumulative case study from 2003 onwards of a select number of information products and services provided by GreyNet. Each were introduced to serve the interests of the grey literature community and their production and supply remain sustained. The first step in the study provides a brief description of GreyNet's types of stakeholders and the nine selected resources intended to serve the grey literature community. Available statistics and data corresponding to the resources are also incorporated here. The second step is the design and implementation of an online survey among GreyNet stakeholders in an effort to determine their involvement. And, in the third and final step an analysis is carried out as to the observed uses of the information resources in step one with the results of the survey conducted among GreyNet stakeholders in step two.

Results from this study seek to establish an accepted level of disparity among the intended purposes, observed uses, and the involvement of stakeholders. The outcome hopes to better enable GreyNet to leverage its information resources – ensuring its positional advantage and enabling it to function more effectively.

Underlying definitions

It is considered of benefit to the reader, when the terms 'leveraging' and 'disparity' are defined early on in the paper, thus ensuring their purposeful meaning throughout the steps carried out in this case study.

Leveraging implies the effective use of a cognitive tool applied in order to improve and/or enhance an organization's positional advantage. It is the power to act effectively on behalf of ones stakeholders by using its key resources to their maximum advantage. For a business this may be interpreted by an increase in profit. For a non-profit organization it may rather imply meeting costs, matching funds, and/or increased use and application of its information products and services.

Disparity implies apparent differences or inconsistencies, when the statistics and data compiled on the key sustained resources do not correlate with the results of the stakeholder survey carried out in the study.

Step One – Identification of Stakeholders and Key Sustained Resources

GreyNet's Business Report 2015[i] served as the primary source for the identification of both the stakeholders as well as the sustained information resources selected in this study. It should be mentioned that the stakeholders are not identified by name but rather by their type of affiliation with GreyNet. Seven types of stakeholders were labelled as follows:

1 – Organizational Member
2 – Recognized Partner
3 – Conference Host and/or Sponsor
4 – Author and/or Researcher
5 – Service Provider
6 – Committee Member
7 – Associate Journal Editor

It should also be mentioned that an individual can be classified in more than one of the seven types of stakeholders indicated above.

Sustained Information Resources

From 2003-2013, nine sustained information resources were selected for this case study. In order to qualify the term 'sustained', each of the selected resources would have had to be at least two years publicly accessible and consecutively accessible from the year of their implementation. The nine sustained resources are as follows:

2003 – Relaunch GreyNet.org[ii]
2003 – Relaunch GL-Conference Series[iii]
2004 – GreyNet Annual Award[iv]
2005 – The Grey Journal, TGJ[v]
2007 – Conference Preprints in OpenSIGLE Repository[1],[2] (later OpenGrey[vi])
2009 – GreyNet Workshop Series[vii,viii]
2010 – GreyNet Social Media[ix]
2012 – GreyNet Datasets[3],[4] in DANS Data Archive[x]
2013 – GreyGuide[5],[6] Repository (and Web Access Portal)[xi]

The first two resources listed above are the GreyNet Website and the TextRelease Website, which powers GreyNet and the GL-Conference Series. In 2003, both of these resources were relaunched after a three year interval. In 2004, the GreyNet Award for outstanding achievement in the field of grey literature became an annual dinner event. And, that following year The Grey Journal (TGJ) was launched as the flagship journal for the international grey literature community. In 2007, GreyNet became open access compliant by making its comprehensive collection of full-text conference preprints online accessible via the OpenSIGLE Repository – later to be renamed OpenGrey. In 2009, GreyNet offered its first summer workshop series, which was later joined by the GreyForum Series – workshops that focus on special topics and interest groups in grey literature. In 2010, GreyNet expanded further its presence via social media by initiating a LinkedIn Discussion group followed more recently by Twitter and Netvibes. In 2012, GreyNet became actively involved in the collection and archiving of its data by way of a project on enhanced publication, whereby data sets were cross-linked to their corresponding conference preprints. And in 2013, the GreyGuide Repository of good practices in grey literature was launched, later to become GreyNet's web access portal.

Sources of Data and Statistics

Available data and statistics upon which this paper is based are drawn from a number of sources, namely • GreyNet's In-house Excel and Outlook files • Conference Evaluation Forms • Web Stats (Network Solutions) • Licensed Journal Stats (EBSCO) • Service Provider Stats (OpenGrey, DANS, and GreyGuide) • and from GreyNet's Social Media (LinkedIn and Twitter).

In the remainder of Step One, current statistics and data that were compiled and which correspond to the sustained information resources selected for this study are presented below. For purposes of comparison, figures from the first two quarters (q1 and q2) of the years 2013, 2014, and 2015 are presented in the tables and their half-year totals are provided in boxes next to each.

---

[1] OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal / D.J. Farace, J. Frantzen, J. Schopfel, C. Stock and N. Henrot. - In: The Grey Journal: An international journal on grey literature. - Amsterdam: TextRelease, Vol. 5, No 1, Spring 2009, pp. 48-52. - ISSN 1574-1796

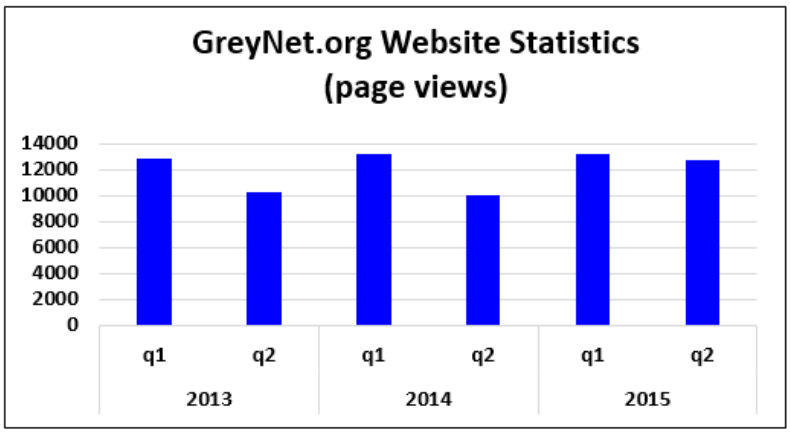
[2] OpenSIGLE – Crossroads for Libraries, Research, and Educational Institutions in the field of Grey Literature / D.J. Farace, J. Frantzen, C. Stock, N. Henrot, and J. Schopfel - In: The Grey Journal: An international journal on grey literature. - Amsterdam: TextRelease, Vol. 6, No 1, Spring 2010, pp. 41-48. - ISSN 1574-1796

[3] Linking full-text Grey Literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012 / Dominic Farace and Jerry Frantzen, Christiane Stock, Laurents Sesink, and Debbie Rabina.– In: GL13 Conference Proceedings, – Amsterdam: TextRelease, 2012. – pp. 143-150. - ISBN 978-90-77484-17-3

[4] Enhanced Publications Project (EPP), Frequently Asked Questions / Dominic Farace, and Jerry Frantzen, GreyNet International, Netherlands; Christiane Stock, Inist-CNRS, France; Paula Witkamp, Data Archiving and Networked Services, DANS, Netherlands. – In: The Grey Journal, Volume 9, No. 2, summer 2013, p. 109. – ISSN 1574-1796

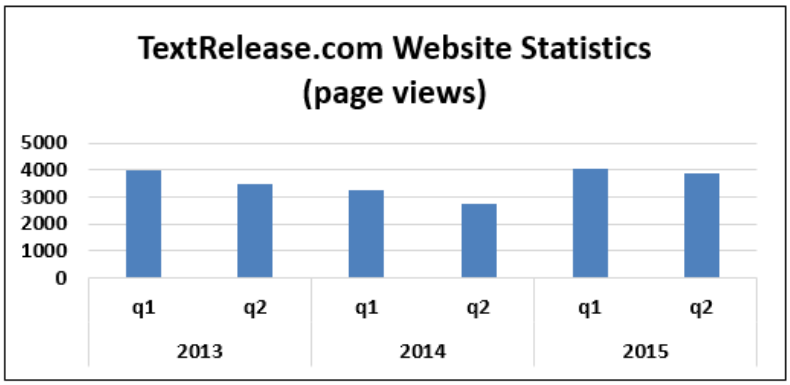
[5] GreyGuide - Guide to Good Practice in Grey Literature: A Community Driven Open Resource Project / Stefania Biagioni, Carlo Carlesi; ISTI-CNR, Joachim Schöpfel; Lille-3, Dominic Farace, Jerry Frantzen; GreyNet International. - In: The Grey Journal, Volume 10, No. 1, Spring 2014, pp. 51-55. – ISSN 1574-1796

[6] GreyGuide, GreyNet's Web Access Portal and Lobby for Change in Grey Literature / Dominic Farace and Jerry Frantzen, GreyNet International; Stefania Biagioni, Carlo Carlesi, and Roberto Ponti, ISTI-CNR; Christiane Stock, Inist-CNRS. - In: The Grey Journal, Volume 11, Special Winter Issue 2015, pp. 25-30. – ISSN 1574-1796



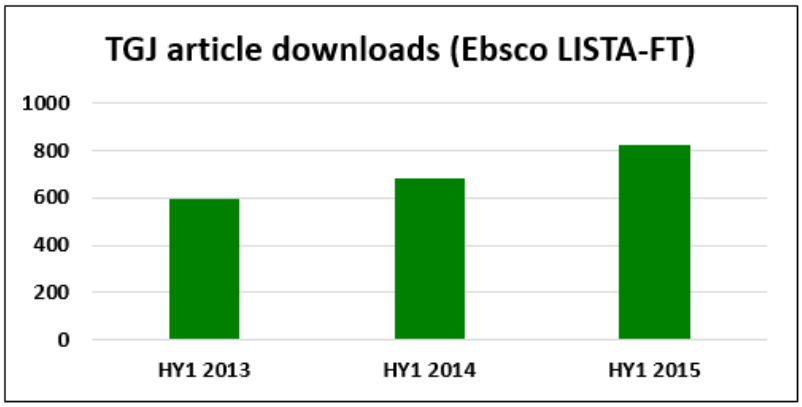
HY1 2013	HY1 2014	HY1 2015
23115	23232	25921

Table 1: Quarterly and half-year totals via Network Solutions, GreyNet's web host



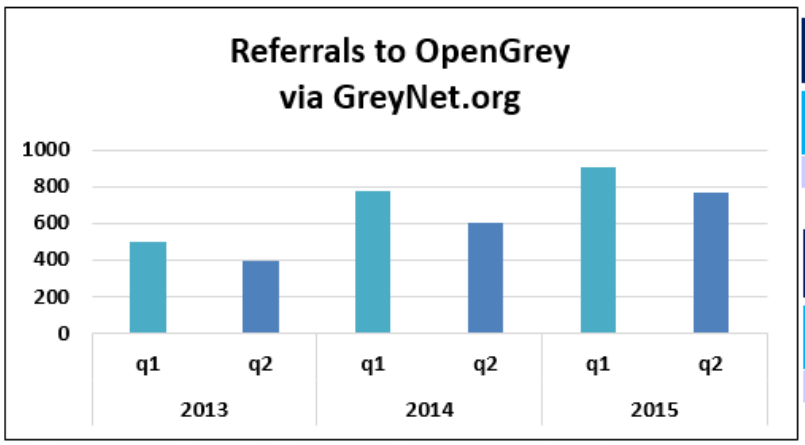
HY1 2013	HY1 2014	HY1 2015
7484	6017	7972

Table 2: Quarterly and half-year totals via Network Solutions, TextRelease's web host



HY1 2013	HY1 2014	HY1 2015
594	680	822

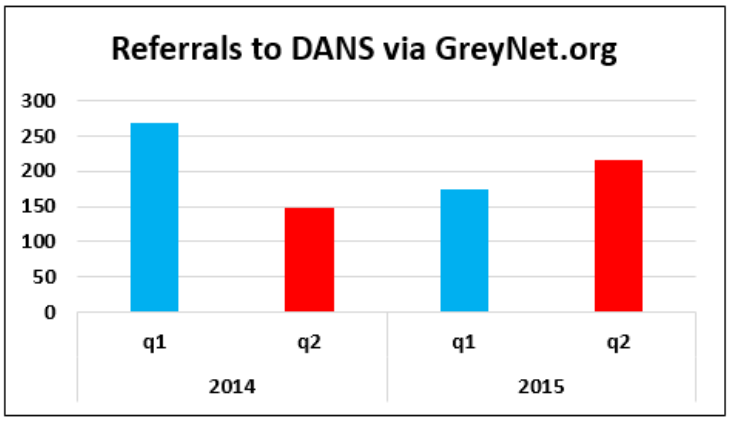
Table 3: Half-year totals via EBSCO, GreyNet's licensing agent



GreyNet Referrals to OpenGrey		
HY1 2013	HY1 2014	HY1 2015
899	1377	1670

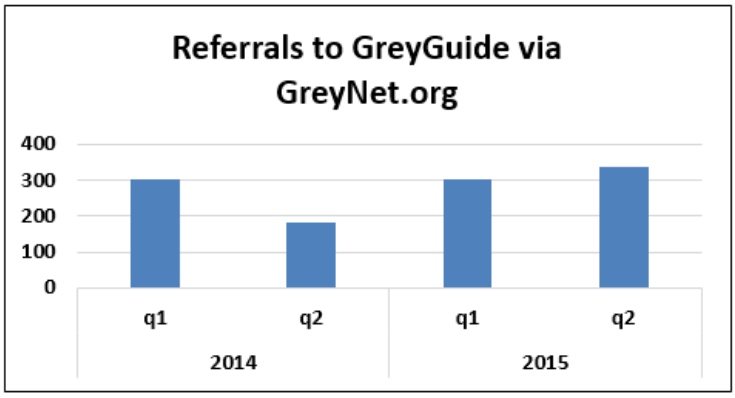
OpenGrey Downloads	
HY1 2014	HY1 2015
210	251

Table 4: Quarterly and half-year totals via GreyNet's website to OpenGrey accompanied by the number half-year downloads of GreyNet preprints and abstracts via the OpenGrey Repository



GreyNet Referrals to DANS	
HY1 2014	HY1 2015
417	390

Table 5: Quarterly and half-year totals via GreyNet's website to the DANS Data Archive



GreyNet Referrals to GreyGuide	
HY1 2014	HY1 2015
487	639

Table 6: Quarterly and half-year totals via GreyNet's website to the GreyGuide Repository

While six of the nine resources in this study could be presented in a similar table format, available data on the other three resources did not easily lend themselves in this way and are thus presented as shown below.

In regard to the **GreyNet Annual Award**, 18 awards have been presented to information professionals working in the field of grey literature in 10 countries worldwide. As to **GreyNet's Workshops**, 7 events have been carried out since 2009 within the GreyWorks Summer Series, and 4 were carried out within the GreyForum Series, which began in 2013. As to **GreyNet's Social Media**, the table below shows available yearly totals:

Social Media	2013	2014	2015
LinkedIn Members	289	530	600
Twitter Followers	-	425	865
Facebook Friends	-	-	65

Step Two – Design and Implementation of an Online Stakeholder Survey

The second step in this study was the design and implementation of an online survey among GreyNet's stakeholders in an effort to determine their involvement and outward satisfaction with the information resources provided. The online questionnaire was carried out over a 12 day period via SurveyMonkey. It was disseminated by way of GreyNet's in-house Distribution List as well as its LinkedIn and Twitter accounts.

There was a total of 77 respondents (100%), 18 (23%) of whom did not identify themselves as a GreyNet stakeholder based on the 7 categories provided. 38% of the respondents chose to enter their email address in the tenth and final question of the survey, while the other 62% remained anonymous. Responses to each of the 10 survey questions varied from 77 (100%) to 62 (81%).

The survey questions and responses appear as follows:

Question 1: Types of GreyNet Stakeholders
GreyNet identifies several types of stakeholders: Members (Associate, Institutional, and Individual), Recognized Partners, Conference Hosts and Sponsors, Authors and Researchers, Service Providers, Committee Members, and Journal Editors

Check the appropriate box (boxes) in which you would place yourself?
(Multiple responses allowed)

Answer Choices	Responses	
Associate, Institutional, or Individual Member	25.97%	20
Recognized Partner	2.60%	2
Conference Host and/or Sponsor	6.49%	5
Author and/or Researcher	51.95%	40
Service Provider	11.69%	9
Committee Member	6.49%	5
Journal Editor	3.90%	3
None of the above	23.38%	18
Total Respondents: 77		

Comments (5)

Question 2: International Conference Series on Grey Literature (ISSN 1385-2308)
Since 1993, sixteen conferences have already taken place in the GL-Series.

In how many of the 16 conferences in this series have you participates?

Answer Choices	Responses	
1-2	27.03%	20
3-4	4.05%	3
5 or more	16.22%	12
None	52.70%	39
Total		74

Comments (0)

Question 3: Annual Award for Outstanding Achievement in the Field of Grey Literature
Over the years, hundreds of presentations have been given in the GL-Conference Series. Each year, the results from written participant evaluations form an important criteria in determining a GreyNet Award recipient.

Approximately, how many GreyNet Award Recipients do you know by name?

Answer Choices	Responses	
1-2	21.43%	15
3-4	8.57%	6
5 or more	17.14%	12
None	52.86%	37
Total		70

Comments (1)

Question 4: The Grey Journal, An International Journal on Grey Literature (ISSN 1574-1796)
The Grey Journal (TGJ) is the only one of its kind in this field of library and information science and for this it received an award in 2008.

Approximately, how many times have you read and/or cited an article from this journal?

Answer Choices	Responses	
1-2	28.99%	20
3-4	21.74%	15
5 or more	17.39%	12
None	31.88%	22
Total		69

Comments (0)

Question 5: GreyNet's Collection of Conference Preprints in OpenGrey
The OpenGrey Repository houses the most current and comprehensive collection of full-text conference preprints in the field of grey literature.

Approximately, how many times have you accessed this repository in the past year?

Answer Choices	Responses	
1-2	35.38%	23
3-4	4.62%	3
5 or more	12.31%	8
None	47.69%	31
Total		65

Question 6: GreyNet's channels of Social Media
Social Media is both a means of connecting communities of shared interest as well as generating new textual and non-textual content. Currently, GreyNet maintains four channels of social media (LinkedIn, Twitter, Netvibes, and its own Listserv).

Approximately, how many times have you posted, shared, or 'liked' content via GreyNet's channels of social media?

Answer Choices	Responses	
1-2	15.38%	10
3-4	6.15%	4
5 or more	10.77%	7
None	67.69%	44
Total		65

Question 7: GreyNet's Published Datasets in DANS Data Archive
Data counts as science output and should be archived for further reference, citation, and use.

Approximately, how many times have you accessed one of GreyNet's datasets either housed in DANS or another data archive?

Answer Choices	Responses	
1-2	6.15%	4
3-4	10.77%	7
5 or more	3.08%	2
None	80.00%	52
Total		65

Question 8: Workshops and Seminars on Grey Literature
The grey literature community spans diverse topics and subject areas and involves organizations in both the public as well as private sector.

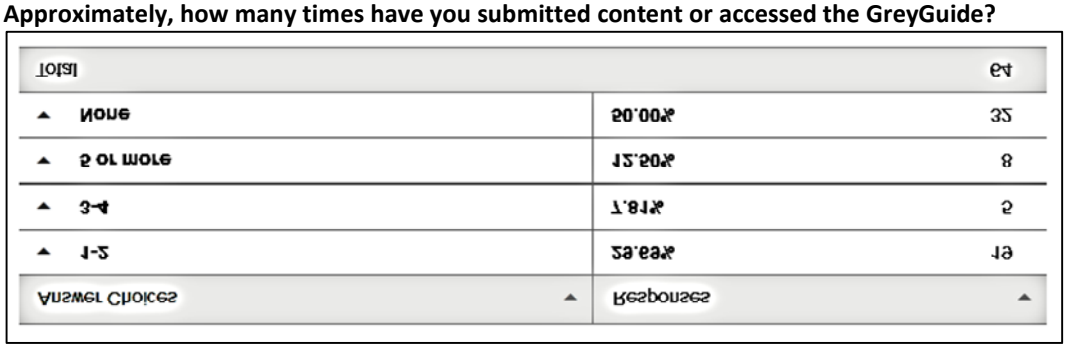
Apart from GreyNet's own workshops and seminars, how many events (onsite or online) have you participated in, where grey literature was featured on the program?

Answer Choices	Responses	
1-2	20.97%	13
3-4	12.90%	8
5 or more	22.58%	14
None	43.55%	27
Total		62

Comments (2)

Question 9: GreyNet's Web Access Portal is the GreyGuide
The GreyGuide Repository and Portal is a community driven enterprise that provides open access to collections and other shared resources in the field of grey literature.

Approximately, how many times have you submitted content or accessed the GreyGuide?



Question 10: Your Name, Organization and Email Address
The results of this Stakeholder Survey will be presented during the Seventeenth International Conference on Grey Literature held on December 1-2, 2015.
If you would be willing to comment further on how GreyNet can better serve its stakeholders, please enter your contact details in the space below.

29 Responses

48 Non-Responses

Step Three – Comparison of the Use of Sustained Resources with the Survey Responses
A comparison of findings from Steps 1 and 2 indicate that data and statistics from five of the nine resources in the study are consistent with responses from at least 50% of the survey respondents. However, data and statistics from the other four resources appear less consistent with responses from the survey as shown in the table below.

Sustained Information Resources	Recorded Data/Statistics	Stakeholder Survey Results
GreyNet.org	↑	**77%**
GL Conference Series	↑	**47%**
GreyNet Annual Award	↑	**47%**
The Grey Journal	↑	**68%**
Conference Preprints in OpenGrey	↑	**52%**
GreyNet Social Media	↑	**32%**
GreyNet Datasets in DANS	↓	**20%**
GreyNet Workshop Series	↑	**56%**
GreyGuide Repository and Web Portal	↑	**50%**

In an attempt to explain the disparity regarding the four resources identified above, we find ourselves tasked with establishing an accepted level of disparity as opposed to an unaccepted level of disparity due to inconsistencies between the statistics/data and the survey responses. It is thought that in so doing, we are then in a position to better identify those information resources that first and foremost require leveraging.

Two resources demonstrating an accepted level of disparity
In Question Two, a near 53% of the survey respondents "never participated" in one of the sixteen GL Conferences, while statistics and data relating to the conference series demonstrate increases. The wording of the question may have been cause for the disparity. Researchers and co-authors, while physically absent from a conference are considered to have participated via their content contribution to the program.

In Question Three, a near 53% of the survey respondents could not name one of the past GreyNet Award Recipients. Here also, the wording of the question may have been cause for disparity. Perhaps the question should have been directed more to the respondent's awareness that an Annual Award for Outstanding Achievement in the field of Grey Literature is presented.

Two resources demonstrating an unaccepted level of disparity
In Question Six, over 67% of the survey respondents had not posted, shared, or liked content via one of GreyNet's channels of social media - while data indicates an increasing number of members and a steady number of new postings. This apparent discrepancy can be explained by the fact that only a relative few stakeholders provide content and feedback via GreyNet's social media. The significant majority can then be categorized as passive recipients. However, they remain potential contributors.

In Question Seven, 80% of the survey respondents had not accessed one of GreyNet's datasets found in DANS or in another data archive. This perhaps can be explained by the fact that only 21 of the 344 full-text conference preprints are currently crosslinked to accompanying data files. Now that the acquisition of data has become part of GreyNet's workflow, there is more potential to increased access and use of the data.

Closing Comments

In order to maintain and improve its positional advantage and ability to act effectively on behalf of its stakeholders, GreyNet will have to explore better ways to leverage its social media as well as the data issuing from the GL Conference Series.

Since the close of the stakeholder survey, a Facebook account has been added to GreyNet's LinkedIn, Twitter, and Netvibes accounts. However this in itself will not ensure that more content and discussion will arise via social media. Likewise since the close of the survey, a poster was

presented at GL17 in order to better inform the authors on how they can enter their research data in the DANS Archive. However, such a demonstration requires a much needed follow-up.

While the results of this study may have produced a method of approach in identifying GreyNet's information resources that require leveraging, the work ahead to accomplish this remains a challenge. A challenge that not only confronts GreyNet, but other grey literature communities[7] as well.

References

[i] http://www.greynet.org/images/GreyNet_Business_Report_2015.pdf
[ii] http://www.greynet.org/
[iii] http://www.textrelease.com/
[iv] http://www.greynet.org/greynetmembership/annualaward.html
[v] http://www.greynet.org/thegreyjournal.html
[vi] http://www.opengrey.eu/search/request?q=greynet
[vii] http://www.greynet.org/researchandeducation/workshops.html
[viii] http://www.greynet.org/greyforumseries.html
[ix] http://www.greynet.org/home/socialmedia.html
[x] https://easy.dans.knaw.nl/ui/?wicket:bookmarkablePage=:nl.knaw.dans.easy.web.search.pages.PublicSearchResultPage&q=greynet
[xi] http://greyguide.isti.cnr.it/

[7] Grey communities: An empirical study on databases and repositories / Helene Prost and Joachim Schopfel. – In: The Grey Journal, Volume 10, Issue 1, 2014, pp. 38-49. – ISSN 1574-1796

the grey Journal

An International Journal on Grey Literature

Subscription Order Form

For GreyNet Members free or reduced rates, see http://www.greynet.org/membership.html

TGJ Volume 12, 2016	*Type of Subscription:*	*Amount in Euros*	*Total*
THE GREY JOURNAL - Printed ISSN 1574-1796 *Annual Subscription, including Postage and Handling*	☐ Institutional	€ 240	**€**
THE GREY JOURNAL - PDF/ CD-Rom ISSN 1574-9320 *Annual Subscription, including Postage and Handling*	☐ Institutional	€ 240	**€**
THE GREY JOURNAL - PDF/ Email / PWP ISSN 1574-180X *Annual Subscription, including Electronic Handling*	☐ Institutional	€ 240	**€**

Customer information

Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
E-mail Address:	

Check one of the boxes below for your Method of Payment:

❑ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to 'TGJ/The Grey Journal'

❑ 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.

Correspondence Address:

TextRelease Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
Tel/Fax +31-(0) 20-331.2420 ● info@textrelease.com ● www.textrelease.com

List of Participating Organisations

African Studies Centre, ASC	Netherlands
Alberta Health Services	Canada
Amsterdam Informatie Netwerk, AIN	Netherlands
Biblioteca Centrale, G. Marconi; CNR	Italy
Boekmanstichting	Netherlands
Centre National de Recherché Scientifique, CNRS	France
Copyright Clearance Center	United States
Data Archiving and Networked Services, DANS-KNAW	Netherlands
Delta Progetti 2000 S.r.l.	Italy
Discoverygarden Inc.	Canada
EBSCO	United States
Eller College of Management, University of Arizona	United States
ETT Solutions	Italy
European Organization for Nuclear Research, CERN	Switzerland
Federal Judicial Center, FJC	United States
Federal Library Information Network, FEDLINK	United States
GERiiCO laboratory	France
German National Library of Science and Technology, TIB	Germany
Grey Literature Network Service, GreyNet International	Netherlands
Hoge Raad der Nederlanden	Netherlands
Hokkaido University	Japan
InfoEthicsUK	United Kingdom
Institut de l'Information Scientifique et Technique, Inist-CNRS	France
Institute of Computational Linguistics, ILC-CNR	Italy
Institute of Information Science and Technologies, ISTI-CNR	Italy
International Council for Scientific & Technical Information, ICSTI	France
Irvine Valley College	United States
Korea Institute of Science & Technology Information, KISTI	Korea
Library of Congress, LoC	United States
Makerere University	Uganda
National Library of Technology, NTK	Czech Republic
National Research Council, CNR	Italy
Network and Information System Office, CNR	Italy
New York Academy of Medicine, NYAM	United States
Nuclear Information Section; International Atomic Energy Agency, NIS-IAEA	Austria
PricewaterhouseCoopers, PwC	Netherlands
ProQuest	United Kingdom
Research Institute on Sustainable Economic Growth, IRCrES-CNR	Italy
Royal Netherlands Academy of Arts and Sciences, KNAW	Netherlands
RSScockpit	Netherlands
Slovak Centre of Scientific and Technical Information, CVTISR	Slovakia
TextRelease, Program and Conference Bureau	Netherlands
University of Belgrade	Serbia
University of Calgary	Canada
University of California, Irvine Libraries, UCI	United States
University Constantine 2	Algeria
University of Illinois at Urbana-Champaign, UIUC	United States
University of Lille 3	France
University of Ljubljana	Slovenia
University of Toronto; Institute of Health Policy, Management and Evaluation	Canada
University of Washington	United States
University of Wisconsin, Milwaukee	United States
University of Zimbabwe	Zimbabwe
Zuckerman College of Public Health, University of Arizona	United States

Eighteenth International Conference on Grey Literature

Leveraging Diversity in Grey Literature

THE NEW YORK ACADEMY OF MEDICINE
HEALTHY CITIES. BETTER LIVES.

1216 Fifth Avenue
New York, NY USA
November 28-29, 2016

Conference Announcement

Scientific information, much of which is published as grey literature, can play a pivotal role in the search for solutions to global problems. Diversity invigorates problem solving and science benefits from a community that approaches problems in a variety of creative ways. Despite their diversity, the hundreds of authors and researchers across the globe involved in grey literature can be seen as part of the same community contributing to the scientific enterprise in valuable ways.

Diversity speaks directly to the effectiveness of information professionals working together as a team and is an essential ingredient for innovation. People from different backgrounds bring with them new information. If you want to build teams, communities, and organizations capable of innovating, you need diversity. It enhances creativity and encourages the search for new information and nuanced perspectives, leading to better decision making and problem solving. Diversity can improve the bottom line of companies as well as organizations, because exposure to it changes the way one thinks. A diverse community of researchers anticipate differences and understand that they will have to work harder to achieve consensus, but their diligence can lead to better outcomes. Authors in the GL-Conference Series come from different societal cultures and geographic regions; however in their research, they are united by the culture of science, which is without borders. This diverse community has over the past two decades applied research methods and offered explanations that have helped this field of information through blind spots, shedding light on what were once seen only as inherent problems. Their evidence based approaches have opened up new areas of research in grey literature. Where in the early '90s the focus was primarily on the demand side of grey literature, equal emphasis today is directed to its supply side. Speed and scale of communication are significant factors that contribute to diversity. The proliferation of technologies has allowed for an exponential growth of knowledge in information science just as in other sciences. However, the diverseness of grey literature resources has become a major challenge to its exploitation. The availability of systems for collecting and aggregating data and its semantic analysis has now become a priority.

GL18 will focus on evidence and will seek to further raise awareness among the wider public to the strength of grey literature based on a shared commitment by a diverse community of authors and researchers responsible for its production and open access. GL18 welcomes information professionals in government and non-government, in academics, business and industry to engage in this year's international forum by responding to the Call for Papers, http://www.textrelease.com/gl18callforpapers.html.

Conference Dateline 2016

● April 15	● May 9	● May 16	● October 15	● November 1	● November 14	● Nov. 28-29
Close, Call for Papers	Author Notifications	Open, Call for Posters	Close, Early Conference Registration	Close, Call for Posters	Submission Conference Papers	GL18 Day One GL18 Day Two Poster Award

TextRelease
GL18 Program and Conference Bureau

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

Eighteenth International Conference on Grey Literature

Leveraging Diversity in Grey Literature

THE NEW YORK ACADEMY OF MEDICINE
HEALTHY CITIES. BETTER LIVES.

1216 Fifth Avenue
New York, NY USA
November 28-29, 2016

Call for Papers

Title of Paper:	**Conference Topic(s):**
Author Name(s):	**Phone:**
Organization(s):	**Fax:**
Postal Address:	**Email:**
Postal/Zip Code – City – Country:	**URL:**

Guidelines for Abstracts

Participants who seek to present a paper at GL18 are invited to submit an English language 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, conference 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

- ☐ **Diversity in Publishing and Disseminating Grey Literature**
- ☐ **Open Access and Shared Grey Resources**
- ☐ **Aggregating Data and its Semantic Analysis**
- ☐ **Public Awareness to Grey Literature via Social Media**
- ☐ **Effectiveness of Grey Literature in Decision Making and Problem Solving**
- ☐ **Other Conference Topic:**

Due Date and Format for Submission

Abstracts in MS Word must be emailed to conference@textrelease.com on or before **April 15th 2016**. The author will receive verification upon its receipt. In early May, shortly after the Program Committee meets, 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
GL18 Program and Conference Bureau

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

Author Information

Aloia, Danielle **75**

Danielle Aloia is Special Projects Librarian at The New York Academy of Medicine. She 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 since 2006, first at AARP and then at the United States. Dept. of Transportation. For the past 4 years she has been managing the Grey Literature Report in Public Health, produced by NYAM. Email: daloia@nyam.org

Basili, Carla **69**

Carla Basili is Senior Researcher at the Italian National Research Council. She has been Associate Professor of Documentation at the Macerata University (1995-2007) and the Lumsa University in Rome (1998-2008), Vice-president of the Italian Association for Advanced Documentation (1998-2004) and Italian delegate in the European Council of Information Associations (ECIA) (1997-2004). Carla Basili is co-ordinator and initiator (since 2001) of the European Network on Information Literacy (EnIL) and of the European Observatory on Information Literacy Policies and Research (since 2006). Since 2009, Carla Basili is the scientific responsible of the research project Information Policies in Science: Knowledge Sharing and Transfer in Scholarly Disciplines. Her research interests focus on Knowledge Sharing and Transfer in the Humanities, Scientific Information Policies and Information Literacy. Selected authored books: La biblioteca in rete: strategie e servizi nella Società dell'Informazione (1998), La biblioteca virtuale: accesso a risorse informative in rete (1994) (with C. Pettenati). Selected edited books: Information literacy in Europe. A first insight into the state of the art of information literacy in the EU (2003), Information Literacy at the crossroad of Education and Information Policies in Europe (2008), The Observatory on Information Literacy Policies and Research in Europe (2009) and Sinergie invisibili. Ricerca e Informazione Scientifica nell'Economia della Conoscenza (2010), Information Policies in the Humanities(2014). Email: c.basili@ceris.cnr.it

Biagioni, Stefania **137, 165**

Stefania Biagioni graduated in Italian Language and Literature at the University of Pisa and specialized in Data Processing and DBMS. She is currently a member of the research staff at the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), an institute of the Italian National Research Council (CNR) located in Pisa. She is head librarian of the Multidisciplinary Library of the CNR Campus in Pisa 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/OpenAire 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. Since 2013 she is involved on the GreyGuide Project. Email: stefania.biagioni@isti.cnr.it

Bouderbane, Azzedine **125**

Azzedine Bouderbane is professeur in library science at the University Constantine 2, Algeria. He holds A Ph.D, a Magister and a Higher Degree in the same discipline. He is fluent In three languages Arabic, French and English. He presented more than 60 papers in national and international conferences. He also published more than 30 papers In scientific journals and collective books. Expert with various international scientific institutions, he is also involved in various scientific research projects. Email: bouderbane.azzedine@yahoo.fr

Cancedda, Flavia **95**

Flavia Cancedda, Technologist, librarian and information specialist. Head of the Italian ISSN Centre (Central Library CNR); Component of "Documentation and Information" Committee of the UNI, National standardization organism (Chairman of the Subcommittee "Presentation, identification and description of documents"; component of the Working Group on National Standard about Professional Requirements for Librarians); Component of the ACNP Libraries Committee (Archivio Collettivo Nazionale Periodici) and of the ACNP Cataloguing Rules Revision Working Group. Since 1991 to 2000 librarian in two different Italian universities and in the National Library of Florence (section: National Bibliography); since 2001 librarian in the Central Library of the Italian National Research Council (CNR). Master's degree in Libraries Management and Direction. Post-graduate degree as Manuscripts conservator at the Special School for Archivists and Librarians (La Sapienza University in Rome). Author of books and essays about the history of libraries and historical bibliography. Forthcoming publications (2015): ISSN e i suoi fratelli. La lunga marcia dei codici internazionali (part 1 and 2), AIB studi Recent publications/contributes: UNI 11535: il riconoscimento che mancava, AIB studi (Online), 55 (2015), 1, 115-124; UNI 11535:2014, Qualificazione delle professioni per il trattamento di dati e documenti - Figura professionale del bibliotecario - Requisiti di conoscenza, abilità e competenza; La norma UNI sulla professione del bibliotecario, Il Mondo degli archivi (Online), feb. 2014; Authority control, workshop, held in FSR, Faster Smarter and Richer. Reshaping the library catalogue, International meeting, Rome, feb. 2014; UNI ISO 8:2014 Documentazione e informazione - Presentazione dei periodici (translation of: ISO 8:1977 Documentation -- Presentation of periodicals). Email: flavia.cancedda@cnr.it

Carlesi, Carlo **165**

Carlo Carlesi, graduated in Computer Science, worked since 1970 at the IEI (now ISTI) of the Italian National Research Council in Pisa. His interest are focused in many software engineering field such as: Development of data base systems, Software quality and testing, Administration and UNIX system management, Digital library systems, Network security and management. In the year 2000-2010 he was head of the "Information Technology Security Service" of the Institute. He participated in many national and international projects, the last being: Italian Project for Research in the Antartic - (South Pole project), the aim of the project was the creation of a Multidisciplinary Integrated Information System to manage and query the Set of Antarctic Data Bases; ERCIM Technical Reference Digital Library - (ETRDL Project), The Digital Library service allows public access through Internet to the technical reports and other not

published document produced by several organizations. It is currently a Research Associate of the Institute ISTI and he is involved in the following projects: PUMA - Publication Management. The Digital Library service allows public access (when permitted) through Internet to the published documents produced by CNR Organizations. GreyGuide - Guide to Good Practices and Resources in Grey Literature. An online forum and repository of good practice and resources in the field of grey literature. Email: carlo.carlesi@isti.cnr.it

Chamberlain Kritikos, Katie **46**

Katie Chamberlain Kritikos is a PhD student at the University of Wisconsin-Milwaukee School of Information Studies. She graduated from the University of Illinois at Urbana-Champaign with a JD (2009) and MLIS (2010) and received her BA, summa cum laude, in English (2006) from the University of Alabama. Formerly, Katie managed the Law Office of Katherine Chamberlain in Princeton, Illinois. Her law practice focused on estate planning, probate, and real estate transactions. With her background in the law and library science, Katie now researches the relationship between libraries, law, and policy. She has published articles on constitutional law and copyright law and is also a certified yoga instructor. Email: kritikos@uwm.edu

Choi, Kiseok **113**

Kiseok Choi is a principal researcher of NTIS Center, Korea Institute of Science and Technology Information. He received the B.S. degree from the school of Computer Science and Statistics at Seoul National University, in 1988, and, M.S. degree from the school of Information and Communication Engineering at Korea Advanced Institute of Science and Technology, in 1997, PH.D. degree from the school of Computer Engineering at Chungnam National University 2013, respectly. His research interests include information retrieval, database and distributed data interoperability. choi@kisti.re.kr

Ćirković, Snježana **131**

Snjezana Cirkovic is Director of the Head Office of the Austrian Academic Library Consortium (Kooperation E-Medien Österreich). Her main responsibilities are organizing and coordinating the license agreements for electronic resources, centralized licensing for all consortia members, negotiating with suppliers, representing the consortium at international level. She holds a Master Degree in Library and Information Science from University of Belgrade, Faculty of Philology, with specialization in private libraries in the 18th century. She is currently writing PhD Dissertation with a specialization in European Cultural History during the 18th and 19th century (German-Serbian Cultural Relationships). From 2004 till 2015 she was working as a Senior Chief Librarian at the Institute for German Studies, Faculty of Philology, Belgrade University. She has developed and taught research courses of Information Literacy and Academic Writing for both M.A. and PhD students at the Faculty of Philology. Her main interests are electronic resources, open access, academic writing and scientific research. Email: snjezana.cirkovic@gmail.com

De Biagi, Luisa **69, 95**

Luisa De Biagi got her Laurea Degree in Literature and Philosophy at 'La Sapienza' Univ. of Rome (Art history and Cultural heritage). With a Specialization in 'Archivist-Palaeographer' (Vatican School of Palaeography, Diplomatics and Archivistics at the Vatican Secret Archive) as well as a Specialization Degree in Archivistics, Palaeography and Diplomatics (Archivio di Stato, Rome) and a Degree from the Vatican School of Library Sciences. De Biagi further holds a Master in 'Business Publishing' (LUISS Management – Rome). She's been working for the SIGLE Network (System for Information on Grey Literature in Europe) since 2002, Since 2010 she's is responsible for the Italian National Referring Centre of Grey Literature at CNR Central Library 'G. Marconi' as representative of the European Network and Openarchive OpenGrey. She's taken part in 3 editions of GreyNet's GL Conference Series (GL5 Amsterdam, GL13 Washington DC, GL14 Rome and GL15 Bratislava). She's is also a member of the CNR Working Group for Cedefop-Refernet Project (Consortium for Professional Education and Training coordinated by ISFOL), the Committee for Legal Deposit Acquisition at CNR Central Library, and a member of the European Association of Health Information and Libraries (EAHIL). She's also responsible for the Library Functional Units 'Education and Training' and 'Cultural Activities Management', organizing didactics laboratories for students, professional training courses and teaching in professional trainings for librarians, students and users. Email: luisa.debiagi@cnr.it

Farace, Dominic **165**

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). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked since 1976. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (SWIDOC/KNAW), Farace founded GreyNet, Grey Literature Network Service in 1992. He has since been responsible for the International Conference Series on Grey Literature (1993-2013). 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: info@greynet.org

Frantzen, Jerry **165**

Jerry Frantzen graduated in 1999 from the Amsterdam University of Applied Sciences/Hogeschool van Amsterdam (AUAS/HvA) in Library and Information Science. Frantzen is the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet, Grey Literature Network Service, as a freelance technical consultant. Email: info@greynet.org

Gelfand, Julia **61**
Julia Gelfand is the Applied Sciences, Engineering & Public Health Librarian at the University of California, Irvine Libraries where over the last 35 years she has performed many roles. She is active professionally and currently is a member of the Board of Directors of the Association of College & Research Libraries (ACRL), a division of the American Library Association, a member of the Science & Technology Section of the International Federation of Library Associations (IFLA) and is Secretary of Section T of the American Association for the Advancement of Science (AAAS). She writes and presents frequently on topics related to Scholarly Communication, Collection Management, Digital Scholarship, integration of multimedia in scientific literature, grey literature, social media, library as publisher. A graduate of Goucher College with graduate degrees from Case Western Reserve University, she is the recipient of many awards including the first GreyNet Award presented in 1999 and has been a Fulbright Fellow and a Thomas J. Watson Fellow. Email: jgelfand@uci.edu

Giannini, Silvia **137**
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. silvia.giannini@isti.cnr.it

Goggi, Sara **104, 137**
Sara Goggi is a technologist at the Institute of Computational Linguistics "Antonio Zampolli" of the Italian National Research Council(CNR-ILC) in Pisa. 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 preminent 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 and since 2011 - her first publication at GL13 - she is working on topics related with Grey Literature. Email: sara.goggi@ilc.cnr.it

Isomoto, Yoshio **44**
Yoshio Isomoto is a librarian of Hokkaido University Library. He graduated from Faculty of Letters, Hokkaido University in 1999 and Faculty of Library, Information Science, University of Library and Information Science in 2001. He has worked in Hokkaido University as librarian since 2001. E-mail: isomoto@lib.hokudai.ac.jp

Južnič, Primož **15**
Primož Južnič is an associate professor at the Department of Library and Information Science and Book Studies at Faculty of Arts, University of Ljubljana (Slovenia). His main area of research and interest is bibliometrics, collection management and LIS education. He teaches the following courses: Bibliometrics, Special libraries, and Collection Management. Before starting his university career, he was a heading different special and academic libraries and information/computer centres. He was also working at the European Commission, for three years, as the seconded informatics expert. primoz.juznic@ff.uni-lj.si

Leroy, Gondy **83**
Gondy Leroy is associate professor in the Department of Management Information at the University of Arizona. She earned a B.S. and M.S. in Experimental Psychology (1996) at the Catholic University of Leuven, Belgium, and a M.S. and Ph.D. in Management Information Systems (2003) at the University of Arizona's MIS department. Her research focuses on NLP in medical informatics and digital government and has been funded by the U.S. National Institutes of Health, National Science Foundation, Microsoft Research and several foundations. She authored the book "Designing User Studies in Informatics", published by Springer, and conducts tutorials on this topic. She is an active advocate for increasing diversity in computing.
Email: gondyleroy@email.arizona.edu

Lin, Anthony **61**
Anthony Lin is the Head of Instruction and Collections at the Irvine Valley College Library. He holds a MSI from the University of Michigan-Ann Arbor, a BA in Spanish from California State University San Marcos, and a BS in Finance from San Diego State University. His interests are emerging technologies, effective bibliographic instruction, and collections management. Email: alin@ivc.edu

Lipinski, Tomas A. **46**
Professor Tomas A. Lipinski completed his Juris Doctor (J.D.) from Marquette University Law School, Milwaukee, Wisconsin, received the Master of Laws (LL.M.) from The John Marshall Law School, Chicago, Illinois, and the Ph.D. from the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Dr. Lipinski has worked in a variety of legal settings including the private, public and non-profit sectors. He is the author of numerous articles and book chapters; his monographs include, *The Library's Legal Answer Book* co-authored with Mary Minow (2003); *The Copyright Law In The Distance Education Classroom (2005), The Complete Copyright Liability Handbook For Librarians And Educators (2006), And The Librarian's Legal Companion For Licensing Information Resources And Services (2012*). Recent articles and chapters include, Click Here to Cloud: End User Issues in Cloud Computing Terms of Service Agreements, in *Challenges Of Information Management Beyond The Cloud: 4th International Symposium On Information Management In A Changing World, Imcw 2013* (Revised Selected Papers.), with Kathrine Henderson, Hate Speech: Legal and Philosophical Aspects, in *The Handbook Of Intellectual Freedom Concepts* (2014), in 2013 with Andrea Copeland, Look before you License: The Use of Public Sharing Websites in building Patron Initiated Public Library Repositories, *Preservation, Digital Technology & Culture* and in 2012, Law vs. Ethics, Conflict and Contrast in Laws Affecting the Role of Libraries, Schools and other Information Intermediaries, *Journal Of Information Ethics.* He has been a visiting professor in summers at the University of Pretoria-School of Information Technology

(Pretoria, South Africa) and at the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Professor Lipinski was the first named member of the Global Law Faculty, Faculty of Law, University of Leuven (Katholieke Universiteit Leuven), Belgium, in Fall of 2006 where he continues to lecture annually at its Centers for Intellectual Property Rights and Interdisciplinary Center for Law and ICT. He is active in copyright education and policy-making, chairing the ACRL Copyright Discussion Group, a member of the ALA OITP Committee on Legislation Copyright Subcommittee, a member of the Copyright and Other Legal Matters Committee of IFLA and serves as an IFLA delegate to the World Intellectual Property Organization's Standing Committee on Copyright and Other Rights. In October of 2014 he returned to the University of Wisconsin—Milwaukee to serve as Professor and Dean of its i-School, the School of Information Studies. tlipinsk@uwm.edu

Marín Arraiza, Paloma **39, 119**
Paloma Marín Arraiza studied physics at the University of Granada (Spain). She also holds a master's degree in information science and science communication from the same university. She developed part of the master's degree at the São Paulo State University (UNESP) in Brazil. Currently, she is part of the Competence Centre for non-textual Materials at the German National Library of Science and Technology (TIB), where she works as a research assistant in the development and research division. Her main interests and research topics are Linked Open Data and videos as a way to communicate science. Email: paloma.marin@tib.uni-hannover.de

Mlambo, Elizabeth **101**
Elizabeth Mlambo is Sub Librarian at the College of Health Sciences at the University of Zimbabwe. She is currently working on a digitization programme funded by DFID and coordinated by British Library Development Studies. This project is a result of bilateral discussions between the British Library Development Studies (BLDS) at the University of Sussex and the University of Zimbabwe. The aim of the project is to improve visibility of social sciences research in particular of that published in developing countries, both current and historical which often remains unseen by an international audience who could benefit from a wider range of perspectives.
Email: emlambo@uzlib.uz.ac.zw

Monachini, Monica **104**
Monica Monachini is a Senior Researcher at CNR-ILC. Field of expertise: computational linguistics, computational lexicography, semantics, lexical semantics, language resources, ontologies, lexicon, terminologies, metadata, validation, methods for retrieving information in different areas (biology, environment, civil protection, oceanography, social media, humanities and social sciences, ...), infrastructural issues related to language resources. Active in many standardisation activities for harmonising lexical information. Involved and responsible of the Pisa team in many international projects for language engineering. Over the last years, she has published articles in the field of lexical resources and information extraction in different areas. Currently, she focused her activities on digital humanities. Member of various Scientific Committees; UNI delegate for ISO/TC37/SC4. Email: Monica.Monachini@ilc.cnr.it

Motta, Gabriele **154**
Gabriele Motta graduated from the "Federico II" University of Naples with a degree in Computer Engineering and specialized in "Ideas, Methods and Applications of Informatics", he currently works at the National Research Council of Italy (CNR) - Network and Information System Office. His main areas of interest are computer system engineering, digital preservation, technological innovation with a focus on public administration. For many years, he had collaborated with the Presidency of the Council of Ministers - Department for Digitization of Public Administration and Technological Innovation – in activities related to public administration modernization, services for citizens and business, dissemination of innovation and e-health.

Mulder, Theo W. **11**
Theo Mulder is a neuropsychologist by training. His primary area of scientific interest is focused on the human motor system (recovery of movement control following brain damage, learning of movements, motor control & cognition, brain-plasticity). He was from 1994 until 1999 professor of rehabilitation research at the Neurological Institute of the Radboud University in Nijmegen. From 1999 until 2006 he was full professor of Human Movement Science and director of the Center for Human Movement Studies at the university of Groningen (University Medical Center Groningen). Since 2006 he is the overarching director of the 16 KNAW institutes in the Netherlands. These institutes range from stem-cell research and brain research to cultural history of the Netherlands, demography, ecology, linguistics and the study of war and genocides. This position brought him in contact with the rich world of the humanities in which he got very much interested. This interest stimulated him to support the further development of the humanities institutes within the KNAW. He is still attached to the University of Groningen as a honorary professor.

Naughton, Robin **75**
Robin Naughton, PhD is the Digital Systems Manager for the New York Academy of Medicine. Prior to joining the Academy, Dr. Naughton was a Digital Consultant with LearningExpress, an EBSCO company, and Oxford University Press, English Language Teaching. She received an Institute of Museum and Library Services (IMLS) fellowship for digital librarianship, and completed her PhD in Information Science from Drexel's College of Computing and Informatics in 2012. Dr. Naughton is a user-centered researcher focused on human-computer interaction (HCI) and mental models, specifically how to design and build interactive systems that are useful, easy to use and enjoyable. Twitter: @robinnaughton

Pardelli, Gabriella **104, 137**
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. 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

Plank, Margret **39**
Margret Plank is currently the Head of the Competence Centre for Non-Textual Materials at the German National Library of Science and Technology in Hannover (Germany). The aim of the Competence Centre for Non-Textual Materials is to develop emerging tools and services that actively support users in the scientific work process enabling non-textual material such as audiovisual media, 3D objects and research data to be published, found and made available on a permanent basis as easily as textual documents. Previously she was responsible for Information Competence and Usability at the TIB. She has also worked as a researcher at the Institute of Information Studies and Language Technology at the University of Hildesheim. She represents TIB on a number of boards including IFLA Steering Committee Audiovisual and Multimedia Section as well as ICSTI / ITOC. Margret Plank holds a Master degree in information science and media studies from the University of Hildesheim, Germany. Email: margret.plank@tib.uni-hannover.de

Prost, Hélène **15**
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: helene.prost@inist.fr

Puccinelli, Roberto **154**
Roberto Puccinelli is currently head of Section I at CNR's "Information System Office" and he's been working for CNR since 2001. He has previously worked in the private sector as system and network engineer. As adjunct professor, he has held courses for the First University of Rome "La Sapienza" ("Operating Systems II") and for the Third University of Rome ("Programming and Computing Laboratory"). He graduated in Electronic Engineering at the University of Rome "La Sapienza" and holds a master cum laude in Enterprise Engineering from the University of Rome "Tor Vergata". In the past he has worked in several research projects in the field of Grid technologies both at the national and international level (executive manager of Work Package 11 within the DataGrid project – V Framework Programme, et al.). He's currently involved in the design and development of CNR's information system. In particular, he coordinates projects for the development of application systems and is responsible for the design and implementation of CNR's data warehouse. He is also responsible for CNR's Local Registration Authority management. He's currently involved in projects regarding the design and development of research product open archives and persistent identifier registers/resolvers. He is author of several articles in the fields of Grid technologies, Autonomic Computing, Software Engineering, Open Archives and Persistent Identifiers. Email: roberto.puccinelli@cnr.it

Reggiani, Lisa **154**
Lisa Reggiani graduated from Sapienza University of Rome with a degree in Classical Studies and specialized in Didactics, she currently works at the National Research Council of Italy - Network and Information System Office and is adjunct professor at the "Guglielmo Marconi" University. For many years she has been involved in research activities in the fields of information and communication sciences, documentation, and learning. She has participated in Italian and European research projects exploring complex phenomena related to the evolution of scientific information and scholarly communication, and of education and training systems. Moreover, she has gained extensive experience in didactic methods and e-learning paths and environments.

Saccone, Massimiliano **154**
Massimiliano Saccone is a researcher at National Research Council of Italy. He works as librarian at the CNR Central Library. He is the responsible for Legal deposit and Open Access activities. He has participated actively in several national and international projects on Information and Knowledge Management (Digital Preservation Europe – DPE, Italian Network of National Bibliography Number – NBN, etc.). His main interests are in scholarly communication, open access, digital preservation, metadata quality control, information system interoperability. Email: massimiliano.saccone@cnr.it

Savić, Dobrica **147**
Dobrica Savic is Head of the Nuclear Information Section (NIS) of the IAEA. He holds a PhD degree from Middlesex University in London, an MPhil degree in Library and Information Science from Loughborough University, UK, an MA in International Relations from the University of Belgrade, Serbia, as well as a Graduate Diploma in Public Administration, Concordia University, Montreal, Canada. He has extensive experience in the management and operations of web, library, information and knowledge management, as well as records management and archives services across various United Nations Agencies, including UNV, UNESCO, World Bank, ICAO, and the IAEA. His main interests are creativity, innovation and use of information technology in library and information services. Email: d.savic@iaea.org

Schöpfel, Joachim **15**
Joachim Schöpfel is senior lecturer at the Department of Information and Library Sciences at the Charles de Gaulle University of Lille 3 and Researcher at the GERiiCO laboratory. He is interested in scientific information, academic publishing, open access, grey literature and eScience. He is a member of GreyNet and euroCRIS. He is also the Director of the National Digitization Centre for PhD Theses (ANRT) in Lille, France.
Email: joachim.schopfel@univ-lille3.fr

Stock, Christiane **165**
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

Index to Authors