

An International Journal on Grey Literature



Spring 2016, TGJ Volume 12, Number 1

‘MINING TEXTUAL AND NON-TEXTUAL DATA SOURCES’

GreyNet

The Grey Journal

An International Journal on Grey Literature

COLOPHON

Journal Editor:

Dr. Dominic J. Farace
Grey Literature Network Service
GreyNet, The Netherlands
journal@grey.net

Associate Editors:

Julia Gelfand
University of California, Irvine
UCI, United States

Dr. Debbie L. Rabina
School of Information and Library Science
Pratt Institute, United States

Dr. Joachim Schöpfel
Université Charles de Gaulle Lille 3
France

Gretta E. Siegel
Portland State University
PSU, United States

Kate Wittenberg
Project Director, Client and Partnership
Development at Ithaka, United States

Technical Editor:

Jerry Frantzen, TextRelease

About TGJ

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



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

CIP

The Grey Journal (TGJ) : An international journal on grey literature / D.J. Farace (Journal Editor); J. Frantzen (Technical Editor) ; GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 12, Number 1, 2016. - *DANS-KNAW (NL), FEDLINK - Library of Congress (USA), CVTISR (SK), EBSCO (USA), Inist-CNRS (FR), ISTI-CNR (IT), KISTI (KR), NIS IAEA (AT), NTK (CZ), and NYAM (USA) are Corporate Authors and Associate Members of GreyNet International.*

ISSN 1574-1796 (Print)
ISSN 1574-180X (PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:

€240 institutional, including postage & handling

Contact Address:

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

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

About GreyNet

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

Full-Text License Agreement

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

© 2016 TextRelease

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

Contents

'MINING TEXTUAL AND NON-TEXTUAL DATA SOURCES'

Move beyond text – How TIB manages the digital assets researchers generate	5
<i>Margret Plank and Paloma Marín Arraiza (Germany)</i>	
Scientific Audiovisual Materials And Linked Open Data: The TIB Perspective	10
<i>Paloma Marín Arraiza (Germany)</i>	
Extracting value from grey literature: processes and technologies for aggregating and analyzing the hidden “big data” treasure of organizations	15
<i>Gabriele Motta, Roberto Puccinelli, Lisa Reggiani, and Massimiliano Saccone (Italy)</i>	
Leveraging Grey Literature – Capitalizing on Value and the Return on Investment: A Cumulative Case Study	26
<i>Dominic Farace and Jerry Frantzen (Netherlands); Stefania Biagioni and Carlo Carlesi (Italy); Christiane Stock (France)</i>	
Geologic Field Trip Guidebooks: Shades of Grey	35
<i>Lura E. Joseph (United States)</i>	
Grey Literature Sources in Historical Perspective : Content Analysis of Handwritten Notes	41
<i>Snježana Čirković (Serbia)</i>	

Colophon.....	2
Editor's Note.....	4
On The Newsfront	
TIB joined GreyNet as an Institutional Member in 2016.....	47
Info Adverts	
GL18 Program Committee – Eighteenth International Conference on Grey Literature.....	48
GL18 Call for Papers – ‘Leveraging Diversity in Grey Literature’	49
FEDLINK, The Federal Library and Information Network - Library of Congress	50
GreyForum 5.1 Wikipedia for Grey Resources: Instruction and Training Skills.....	51
TGJ Subscription Form 2016	54
Author Information.....	52
Notes for Contributors.....	55

EDITOR'S NOTE

Eighteenth International Conference on Grey Literature *Leveraging Diversity in Grey Literature, 28-29 November 2016*

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>.

Dominic Farace,
Journal Editor

Move beyond text – How TIB manages the digital assets researchers generate^{*}

Margret Plank and Paloma Marín Arraiza (Germany)

Abstract

The supply, use and significance of non-textual materials is steadily increasing in the areas of research and teaching. Digital assets like scientific videos, 3D objects and research data are highly relevant in order to make science reproducible. Yet, they often present a lack of appropriate metadata, unique identification, and long-term preservation, remaining beyond the relevant and journal-based scientific publication system. The TIB has established a competence centre for non-textual materials in order to improve the access and the use of those digital assets. As a use case of a service for scientific videos the TIB|AV-Portal is presented. Further best practices are also discussed.

Keywords: non-textual materials, scientific videos, access, multimedia retrieval, library services, scientific communication, infrastructure.

1. Introduction

Today's research papers move beyond text as they include a variety of comprehensive digital assets such as research data, audiovisual media, 3D-objects and even software code. Those digital assets have a unique life cycle within their scientific communities. However, only a negligible proportion of those digital assets are accessible at present, whilst scientific texts are, in principle, sufficiently well-documented and available. This can lead to serious problems when it comes to preservation and potential of re-use. Consequently, valuable scientific information cannot be found, cited or re-used and remain hidden when a research project ends or the responsible scientists transfer to other institutions (Kraft et al., 2015).

In 2011 TIB established a competence centre for non-textual materials, which focuses on strategies regarding "Move beyond text". The competence centre aims at improving ease of access and use of non-textual material, such as audiovisual media, 3D objects and research data. It concerns itself with the development of tools and infrastructure that actively support users in the scientific work processes, enabling the easy publication, finding and long-term availability of non-textual media. To accomplish this, the TIB actively engages in communities, providing support and advice to fellow libraries and other preservers of scientific knowledge and its associated digital assets. This paper addresses how the TIB manages the digital assets researchers generate using the example of scientific audiovisual media. The TIB|AV-Portal¹ shows how to overcome the appearing challenges regarding scientific videos in particular.

The structure of the paper is as follows: this section gives an overview of the situation and presents the problem. Section 2 approaches the management of scientific audiovisual media from a library perspective and presents the TIB|AV-Portal as use case for scientific video service. Section 3 summarises the essential conclusions.

2. Managing scientific audiovisual material in the library context: The TIB|AV-Portal

2.1 Background

Scientific audiovisual media such as computer visualizations, simulations, research laboratory experiments, interviews and recordings of lectures and conferences have become an important part of scientific communication (Spicer 2014). These media contribute to a better understanding and discussion of research work and results because they can describe dynamic phenomena which cannot be precisely detailed in words and pictures (Whitesides for the American Chemical Society, 2011).

The publication of scientific audiovisual media goes beyond the traditional journal-based scientific publication system (Löwe et al., 2015). Audiovisual scientific media are usually shared using web portals that do not contribute to a long-term preservation. There is also a lack of metadata attached to these media (Nixon and Troncy, 2014), which makes the search and retrieval process very difficult. According to these issues, scientific audiovisual media can be considered non-textual grey literature. Their management is a challenge affecting academia (Löwe et al., 2015). Often, scientific videos are uploaded to general video-sharing platforms, such as YouTube, Vimeo or Dailymotion, where videos are provided with little metadata – let alone individual segments. A manual detailed description and annotation of

^{*} First published in the GL17 Conference Proceedings, February 2016.

¹ <https://av.getinfo.de/>

the content is considered too time-consuming (Snoek, 2007). Therefore, there are several initiatives that introduce automated metadata extraction for scientific and non-scientific videos. Neumann and Plank (2013) enumerate two of them: (i) ScienceCinema², the video portal of the Office of Scientific and Technical Information (OSTI) of the U.S. Department of Energy and the European Organization for Nuclear Research (CERN), where audio indexing and speech recognition is used; (ii) Voxlead New³, which presents a navigation within the video based on speech recognition.

Further examples of portals outside the library world are SciveeTV⁴, JOVE⁵ and YoVisto⁶. SciveeTV identifies scientific videos with a Digital Object Identifier (DOI) and offers the possibility of uploading supplementary materials. The portal JOVE presents a subscription-format 'video journal', in which a video is created after the submission of a manuscript. Videos are properly indexed and provided with use metrics. YoVisto presents a portal for academic video search, where the videos are segmented and automatically tagged according to the content of each segment. Innovative semantic searches can be also performed.

With the project The Open Video Digital Library⁷, the library approach to video indexing begins to change. Traditionally, librarians have applied textual bibliographic metadata to videos (Marchionini and Geisler, 2002). However, there is a need of new indexing models that can describe the characteristics of a video. The framework of The Open Video Digital Library includes tasks such as visual and linguistic analysis. This has a direct effect on the way video characteristics are indexed.

Following new ways of indexing and in order to implement innovative video access and retrieving as a library service, the Competence Centre for Non-Textual Materials at TIB teamed up with the Hasso Plattner Institute to develop the TIB|AV Portal.

2.2 The TIB|AV-Portal

The TIB|AV Portal is a web-based platform for scientific videos from the fields of engineering, architecture, chemistry, information technology, mathematics, and physics. The hosted videos mainly consist of computer visualisations, learning material, simulations, experiments, interviews, and recordings of lectures and conferences. Additional materials related to the video can also be uploaded.

The TIB|AV-Portal presents diverse innovative features. Key features are the identification of each hosted video with a DOI, which provides a stable identification for the video, and the use of different kinds of automatic video analysis. These video analyses technologies provide a wide range of metadata associated with the video as well as an audio transcription. Moreover, most videos have been published under a Creative Commons⁸ license.

Scientific audiovisual media require more metadata.

As defined by Guenther and Radebaugh (2004, p-1), metadata are 'structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource'. The information resources in the TIB|AV-Portal are video materials that are described by two types of metadata: manual and automatic.

The manual metadata, which are provided by publishers and video providers, describe the whole video. They are sub-classified in three groups: formal (e.g. title, author or publisher), technical (e.g. file size or duration) and content-related (e.g. subject, abstract/description or keywords). Due to the fact that these metadata are intellectually created, they present a high reliability during the search and retrieval process.

The automatic metadata are generated after a process containing a structural, text, speech, and image analysis. The process consists of five steps. First, the video is split into segments based on the visual content. Thus, each segment becomes an individual element that can be retrieved by the search engine. Secondly, the written language on the video is indexed using Optical Character Recognition (OCR). This allows the searches within the written language. Thirdly, speech to text converts spoken language in the video into a speech transcript. Fourthly, moving images are detected by visual concepts and are indexed according to subject-specific and multidisciplinary visual concepts (Blümel et al, 2012). Finally, named-entity recognition indexes the video segments with semantically associated terms based on the transcripts from OCR and speech recognition (Strobel and Plank, 2014).

² <http://www.osti.gov/sciencecinema/>

³ <http://voxlead.labs.exalead.com/>

⁴ <http://www.scivee.tv/>

⁵ <http://www.jove.com/>

⁶ <http://www.yovisto.com/>

⁷ <http://www.open-video.org/>

⁸ <http://creativecommons.org/>

Fig. 1. Current video⁹ detail page of the TIB|AV-Portal.
Automatic metadata are shown in the right side.

As a result of this process, new metadata are associated with individual video segments, allowing precise searches within the content of the video. Therefore, the searchability of the video is enhanced. Similar approaches are also carried out by publishers such as SAGE Journals or Alexander Street Press (Davies, 2015).

Searching and discovering scientific audiovisual media.

In most cases, the search is only performed within metadata that describe the whole video and not in the individual segments (Strobel and Plank, 2014). There is a need to improve these searches, hence the use of automatic generated metadata, and recognised entities to perform semantic searches in the portal as well as within the video content.

⁹ <https://av.getinfo.de/media/14624?359>

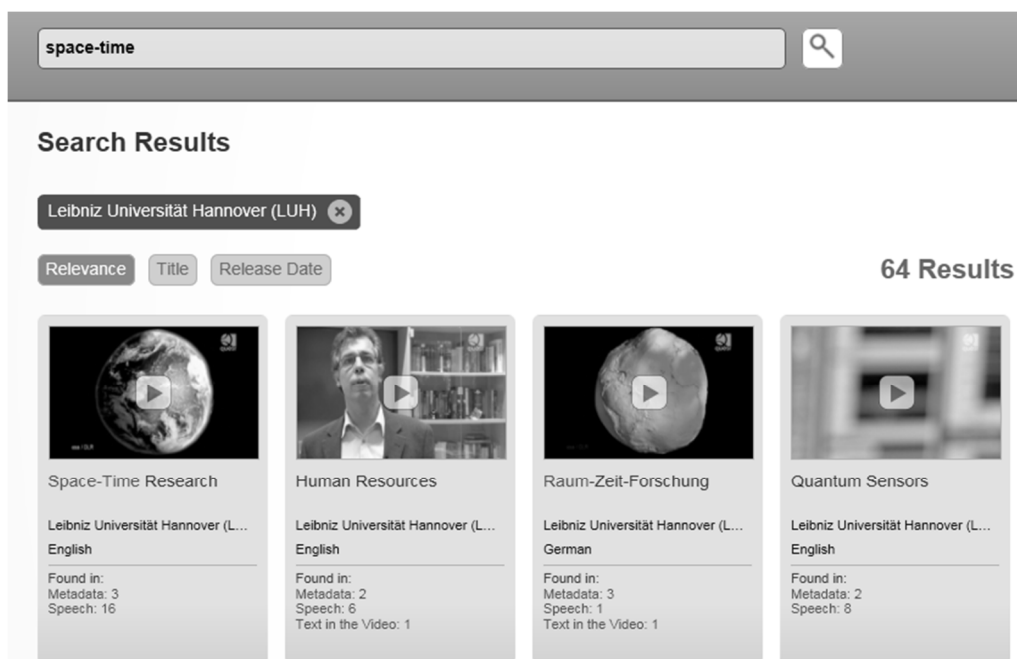


Fig. 2. Search results page of the TIB|AV-Portal.
Searched term matches highlighted in red colour.

As shown in figure 3, the terms in the query can be found in the authoritative metadata like author, title and abstract as well as in the spoken text, text overlays or image information. The terms in the query are also semantically associated with other index terms (synonyms, English-German translations, etc.). This allows the users to discover new videos that otherwise would have remained hidden. For instance, retrieving a video in German when the query was in English (see figure 3).

2.3 Linking scientific content, media, and data

The upload of additional material related to the video complements the information enhancement. Causal associative metadata relationships, such as 'isSupplementBy', 'isSupplementTo', 'cites' or 'isCitedBy', allow the user to connect a video and other related materials. In order to cover the wide spectrum of these materials, different file formats (doc, .cdv, .odt, .pdf, .jpeg, etc.) are accepted in the portal.

Metadata			
Formal Metadata		Content Metadata	
Title	Soziale Netzwerke für Forschende und akademisches Identitätsmanagement	Subject Area	Other
Title of Series	CoScience Open Video Lectures	Accompanying Materials	
Part Number	5		
Author	Peters, Isabella Leiß, Caroline Hohmann, Tina Heller, Lambert	Kapitel Soziale Netzwerke für Forschende im Handbuch CoScience	
Contributors	Mehlig, Martin Schrenk, Philip		
License	CC Attribution 3.0 Germany: You are free to use, adapt and copy, distribute and transmit the work or content in adapted or unchanged form for any legal purpose as long as the work is attributed to the author in the manner specified by the author or licensor.		
DOI	10.5446/15234		
Publisher	Technische Informationsbibliothek Hannover (TIB)		
Release Date	2015		
Language	German		
Technical Metadata			
File Size	2161MB		
Duration	59:08		

Fig. 3. Metadata section of a video¹⁰ at TIB|AV-Portal.

As accompanying materials, not only text files are available, but also links to material in other platforms. The CoScience Handbook¹¹ is shown as instance.

¹⁰

<https://av.getinfo.de/media/15234?338>

Future directions in the management of audiovisual scientific information consist of the publication of the data as Linked Open Data, following W3C Best Practices¹². This will provide a better data interoperability, and reuse.

3 Conclusions

The use of new types of scientific information, beyond the traditional journal-based system, is an ongoing process. Videos, images or research data are part of the research process and as relevant as the final paper. Academic libraries, as information managers, should provide new infrastructures that archive this information and give it an appropriate identification and term of use.

This paper focuses in particular on the treatment of scientific videos from the library perspective. The TIB|AV-Portal shows how videos can be identified and indexed with manual and automatic generated metadata. These metadata contribute to a better description of the video as a whole and at the segment level, facilitating the access and reuse.

It is also important to establish links between video, as non-textual information, and complementary sources of information, denoted in the portal as 'Accompanying materials'. This shows how the portal achieves one of the technical principles of the Pisa Declaration (2014), namely: 'encourage system for linking data and other non-textual content to their grey literature publications together with interoperability standards for sharing grey literature'.

Along this article, we have focused on the importance of new types of information in science. New ways of indexing have been shown using the TIB|AV-Portal as example, and it has been discussed why academic libraries need to provide new information services and infrastructure.

References

1. American Chemical Society. (2011, April 29). Publishing Your Research 101. Impact of video on scientific articles. [Video file]. [Retrieved: September 9, 2015] Available in: <https://www.youtube.com/watch?v=HboNzrqOMKE>
2. Blümel, I., Hentschel, Ch., H. Sack (2013). Automatic Annotation of Scientific Video Material based on Visual Concept Detection. Proceedings of i-KNOW 2013, ACM, 2013, article 16. Available in: <http://dx.doi.org/10.1145/2494188.2494213>
3. Ayuso García, M. D. and Martínez Navarro, V. (2014) La literatura gris en entornos digitales: estrategias de calidad y evaluación. In: Revista Interamericana de Bibliotecología. V. 27, num. 2 (jul.-dec. 2004), pp. 49-70.
4. Davies, N. (2015). Video Enhances Publications. Research Information. [Retrieved: September 9, 2015] Available in: http://www.researchinformation.info/features/feature.php?feature_id=499
5. Guenther, R., Radebaugh, J. (2004) Understanding Metadata. National Information Standards Organization. Bethesda, MD: NISO Press. [Retrieved: September 9, 2015]. Available in: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>.
6. Grey Literature Network Service. Pisa Declaration on Policy Development for Grey Literature Resources. May 16, 2014. [Retrieved: September 9, 2015]. Available in: http://www.greynet.org/images/Pisa_Declaration_May_2014.pdf
7. Kraft, A., Löwe, P., Plank, M., Heller, L., Dreyer, B. (2015). Preserving the long tail in a big data world: Frameworks for e-infrastructures in research libraries. Proceedings 2015 PV Conference, Darmstadt, Germany, 3-5 November 2015. (To be published)
8. Löwe, P., Plank, M., Marín-Arraiza, P. (14th – 17th July 2015) Acquisition of audiovisual Scientific Technical Information from OSGeo by TIB Hannover: A work in progress report. In: Geomatics Workbooks n° 12, FOSS4G Europe Como (Italy).
9. Marchionini, G., Geisler, G. (2002). The Open Video Digital Library. In: D-Lib Magazine, vol. 8, n. 12. Available in: <http://www.dlib.org/dlib/december02/marchionini/12marchionini.html>
10. Neumann, J., Plank, M. (2013). TIB's Portal for audiovisual media: New ways of indexing and retrieval. In: IFLA WLIC 2013. Available in: <http://library.ifla.org/92/1/124-neumann-en.pdf>
11. Nixon, L., Troncy, R. (2014). Survey of Semantic Media Annotation Tools for the Web: Towards new Media Applications with Linked Media. In: 11th ESWC 2014, Demos Track. V.CCIS 476, pp.100-114. Available in: http://dx.doi.org/10.1007/978-3-319-11955-7_9
12. Snoek, C.G.M., Huurnink, B., Hollink, L., de Rijke, M., Schreiber, G., Worring, M. (2007). Adding Semantics to Detectors for Video Retrieval. In: IEEE Trans. Multimedia, V. 9, num. 5, pp. 975-986. Available in: <http://dx.doi.org/10.1109/TMM.2007.900156>.
13. Spicer, S. (2014). Exploring Video Abstracts in Science Journals: An Overview and Case Study. In: Journal of Librarianship and Scholarly Communication V.2, num.2, eP1110. Available in: <http://dx.doi.org/10.7710/2162-3309.1110>
14. Strobel, S., Plank, M. (2014). Semantische Suche nach wissenschaftlichen Videos: Automatische Verschlagwortung durch Named Entity Recognition. In: Zeitschrift für Bibliothekswesen und Bibliographie. V.61, num. 4-5, pp. 255-259. Available in: <http://dx.doi.org/10.3196/18642950146145154>
15. Wittenberg, K. (2008). The Role of the Library in the 21st-Century Scholarly Publishing. In: No Brief Candle: Reconceiving Research Libraries for the 21st Century. Council on Library and Information Resources. [Retrieved: September 9, 2015]. Available in: <http://www.clir.org/pubs/reports/pub142/wittenberg.html>

¹¹ <http://handbuch.io/w/Hauptseite>

¹² <http://www.w3.org/TR/ld-bp/>

Scientific Audiovisual Materials And Linked Open Data: The TIB Perspective^{*}

Paloma Marín Arraiza (Germany)

Abstract

Libraries are starting to use Linked Open Data (LOD) to provide their data (library data) for reuse and to enrich them. However, most initiatives are only available for textual resources, whereas non-textual resources stay aside. Firstly, this paper discusses the potential of library data to be published as LOD. Secondly, it focuses on the library data related to the management of audiovisual scientific materials in the TIB|AV-Portal. The use LOD Standards to support multilingual functionalities and data reuse is outlined. Future developments lead to the building of semantic applications based on LOD Structures

Keywords: scientific videos, linked open data, data publishing, semantic applications, grey literature, library data.

1 Introduction

A large amount of data is generated by global science every day from multiple sources. Libraries' task is to organise and classify all the resources and their respective data, generating new data known as 'library data'.

Data on the web exists in a structured form in databases, and in semi-structured forms in textual and non-textual collections. To deal with both the flood of information as well as the range of heterogeneous data formats, a new approach was needed for information searching and access. The data should no longer be isolated but connected to other data, and become accessible, explorable and discoverable by both people and machines. Thus, the HTML based web changed into the Web of Data, also referred to as Semantic Web. It derives value from data and creates pathways between datasets and resources (Heath and Bizer, 2011).

The derivation of value from the data is possible due to the structure this data presents. In the web of data, information and knowledge are stored in simple structures known as 'triples', which consist of three parts: subject, predicate and object. To publish and interlink structured data on the web, RDF (Resource Description Framework) is used. Data linked to other and published under open licenses (such as CC0¹) are known as Linked Open Data (LOD).

Tim Berners-Lee (2006) outlined the principles structured data need to follow in order to be published as Linked Data:

1. Use URIs (unique resource identifier) as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
4. Include links to other URIs, so that they can discover more things.

The use of semantic technologies and LOD concerns libraries and has a lot of advantages. LOD is better retrievable and easier to interlink since URIs, that define things, are more stable than URLs (uniform resource locator), that define addresses. The interoperability and further use of the data is improved since LOD is based on open web standards and besides the used data model, RDF, is more flexible than other library standards (e.g. MARC) (Pohl and Danowski, 2011). Therefore, W3C Standards become a reliable and flexible alternative.

The aim of this article is to present the use of LOD in libraries and at the TIB, specifically in the TIB|AV-Portal, and how LOD support multilingualism, data reuse and information retrieval applications.

The structure of the paper is as follows. Section 1 introduces the topic and describes the structure of the paper. Section 2 defines library data and specifies the use of LOD in libraries for textual and non-textual materials. Section 3 presents the use of LOD in the TIB|AV-Portal. Section 4 concludes the article.

2 Linked Open Data in Libraries

2.1 Library Data and new types of scientific information

In the Library Data Incubator Group Final Report (2011), library data is defined as 'any type of digital information produced or curated by libraries that describes resources or aids their discovery'. So far, scientific written texts such as books and journal articles were the main resource of academic libraries.

^{*} First published in the GL17 Conference Proceedings, February 2016.

¹ Creative Commons Zero, public domain.

Non-textual materials were considered to be inappropriate for academic purposes, and were usually regarded as general-interest publications rather than as ‘proper’ scientific publications (Löwgren, 2011). However, the scientific landscape changes constantly and academic librarians must be aware of the importance of this multimodal scholarship (Spicer, 2014). This means, they should provide platforms to support these materials, index them properly, and promote their accessibility and reuse.

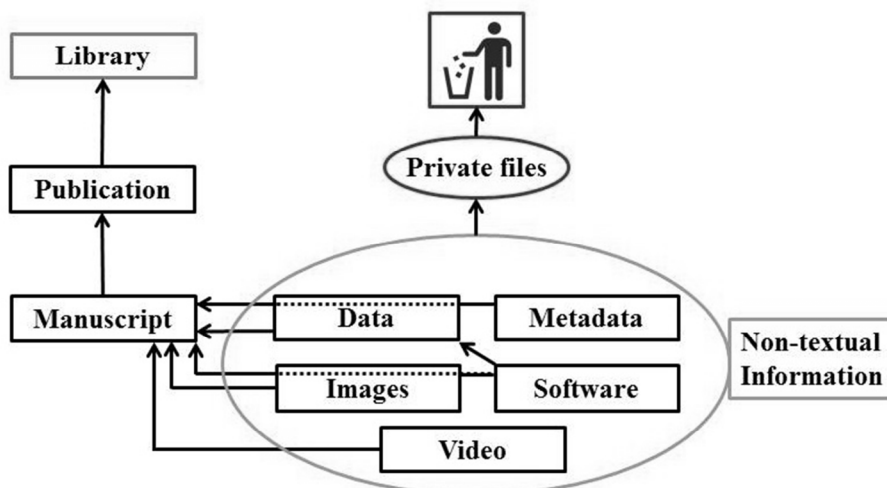


Fig. 1. Flow of scientific information from research to publication (modified after Klump et al., 2006²).

Therefore, as mentioned in the Pisa Declaration (2014), ‘grey materials’ should be persistent identified and linked, as far as possible, with other publications. This generates a big amount of library data that aids the classification and finding of relevant information. Not doing so lead to a loss of scientific information, as shown in figure 1.

2.2 The potential of library data for Linked Open Data

The W3C consider libraries an important focus for LOD because of being important content providers. Being usually structured information, library data present a high potential to be shared as LOD. Peset et al. (2011) highlight libraries projects where LOD play a crucial role: Authorities and vocabularies of the Library of Congress, Linked Data service of the Deutsche Nationalbibliothek (DNB), and Libris, the collaborative Swedish catalog. Moreover, the Web is considered a global data space where libraries should publish their data using RDF and open licenses. Doing so, they promote and support data reuse (Saorín, 2012).

However, most library content is text-based providing few non-textual materials. An example of aggregation environment with textual and non-textual materials is Europeana. Europeana is classified by the Library Data Incubator Group under ‘archives and heterogeneous data’ since it manages metadata from diverse types of materials and formats (Ríos-Hilario et al. 2012). Some tasks performed by Europeana to promote LOD in aggregation scenarios are the experimentation with semantic search based on RDF within collections, knowledge organization systems alignment, and data aggregation system with LOD export.

In this context, we present the TIB|AV-Portal as future use case of the LOD approach for non-textual materials. Some applications, such as multilingualism, are already implemented in the portal. Further implementation is focused on the weaving the data into LOD to enable metadata enrichment in scientific audiovisual materials.

3 The TIB|AV-Portal: Current practices and perspectives with Linked Open Data

3.1 Multilingualism

Textual metadata from text and speech recognition are linked to entities of the *Gemeinsame Normdatei* (GND, German for: Integrated Authority File), which is mainly managed by the German National Library. The data of the GND are available with the formats MARC 21 Authority, MARC21-XML and RDF-XML

² Klump et al. (2006) Data Publication in the Open Access Initiative. Data Science Journal, Volume 5, 15 June, 2006, page: 80.

under Creative Commons Zero³ (CC0). This publishing as Linked Open Data enables the linking between entities in the portal.

From the complete dataset of the GND, the TIB|AV-Portal uses the terms related to the six TIB core subjects. The GND includes synonyms, homonyms, hierarchical relations between terms and cross-references between related terms. These properties are used in the portal to perform a semantic search.

In order to achieve a multilingual and semantic portal, further terms in English were needed, since the GND possessed few English labels. A mapping into other controlled vocabularies was required. For this purpose, labels from the DBpedia⁴, Library of Congress Subject Headings⁵ (LCSH), results of the project 'Multi Lingual Access to Subjects'⁶ (MACS) and the WTI thesaurus 'Technology and Management'⁷ were selected. These terms were also available as LOD and are now saved in the local RDF store (Strobel, 2014).

The combination of vocabularies and bilingual labels enables a cross-lingual search in the portal, as shown in figure 2.

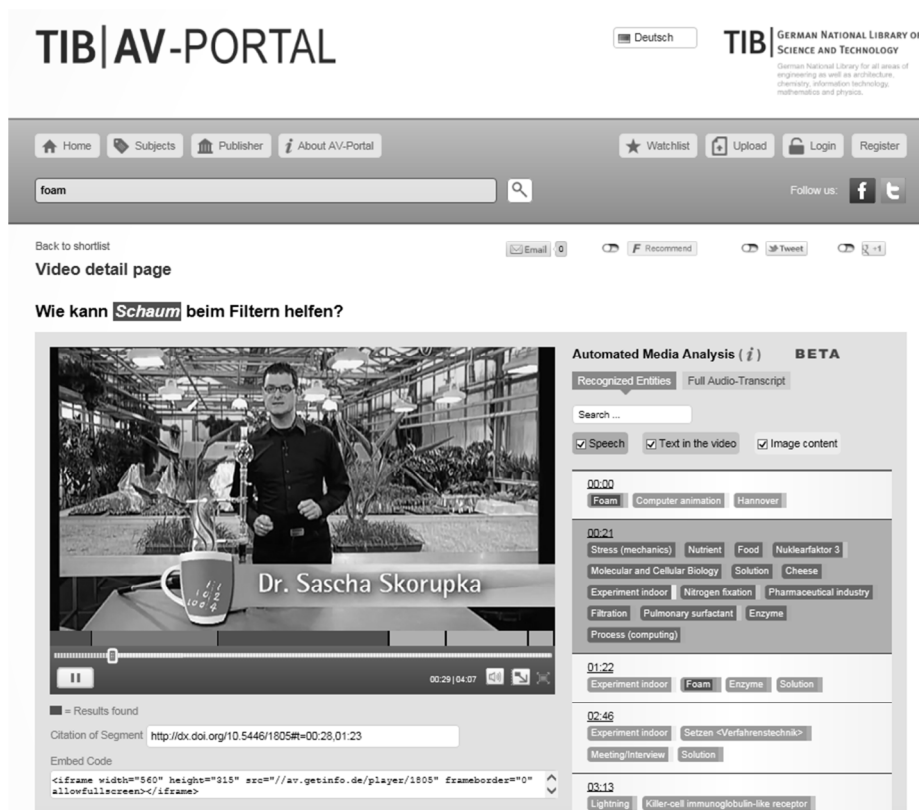


Fig. 2. Video detail page of the TIB|AV-Portal.
Results of the cross-lingual retrieval (foam-Schaum) are shown.

This is likely to increase the completeness of relevant search results and improve video indexing based on LOD properties. The semantic search based on linked data is an interesting research topic in continuous development (Waitelonis and Sack, 2011; 2014).

3.2 Data publishing and semantic applications

A fundamental question is how to publish metadata as LOD. According to Zuiderwijk et al. (2012) there is no standardised way to do it and one finds different approaches. Heath and Bizer (2011) also support this idea and describe possible Linked Data publication patterns. The type of input data (queryable structured data, static structure data or text documents) determines the appropriate publishing pattern. Static structured data coming from XML files must be converted into RDF files directly into an RDF store (Heath and Bizer, 2011). The data from a RDF store (also known as triple store) can be directly published as Linked Data if they fulfill the Linked Data principles.

³ <https://creativecommons.org/publicdomain/zero/1.0>

⁴ <http://de.dbpedia.org>

⁵ <http://id.loc.gov/authorities/subjects.html>

⁶ http://www.dnb.de/DE/Wir/Kooperation/MACS/macs_node.html

⁷ <http://www.wti-frankfurt.de/index.php/produkte-thesaurus>

Due to the structured nature of the data of the TIB|AV-Portal and the existence of an RDF store, the weaving into Linked Data requires the fulfillment of the principles. To do so, parts of the internal ontology and URI scheme need to become dereferenceable. Therefore, we are mapping and merging our internal ontology with existing LOD vocabularies⁸. Following best practices for library data, Dublin Core Terms (dcterms), Schema.org, Bibliographic Framework Initiative (Bibframe) and The Bibliographic Ontology (BIBO) were identified as best vocabularies for our purposes. Dealing with videos and their automatic annotation made also important the use of the Ontology for Media Resources and the Open Annotation Model as part of the mapping process.

The datasets will be exposed as dumps, as already done by other libraries such as The European Library⁹. Dumps are an efficient and fast way to share data with consumers. However, it requires a content and synchronization planning. As data provider, one have to decide which data are part of the dump, its format and actualization frequency. Following other libraries actions, we decide to provide 3 dumps (one with the IWF¹⁰ collection, one with the TIB core subjects collection and one with the whole collection). The chosen formats are RDF application/rdf+xml and RDF in text/turtle, and we will offer a quarterly actualization frequency.

Further, the metadata can be embedded into the portal in form of RDFa. The last specification of RDFa 1.1¹¹ allows the use of RDF in HTML. This provides more control on data access and may be built on top of an RDF store. RDFa defines attributes for the semantic markup such as *about*, *property* or *vocab* to identify the subject, relations between subject and object, and the vocabularies that are used (Strobel and Marín-Arraiza, 2015). We consider RDFa an appropriate form to provide more relevant information –for instance, about an author– and to retrieve information available elsewhere in our portal or on the web –for instance, articles written by the same author and retrieve through the property *schema:author*–. This should establish a net between our content and related content of other providers, supporting reuse and content promotion.

4 Conclusions

Data organization is a requirement that libraries have to cope with. Nowadays, library data results not just after having managed textual information but also non-textual information.

The TIB|AV-Portal manages scientific videos as part of the non-textual information environment and generates a big amount of data associated with each hosted video. The data is indexed according to international standards and W3C Best Practices. However, it is still possible to go a step further and make the data available to third parties.

In other to execute this action, LOD publishing patterns are being followed. This guarantees that our data are no longer isolated or left behind the information world, which, unfortunately, happens often with non-textual information in science. Moreover, the data can be reused by others, and being part of Linked Data Cloud¹² improves the visibility of the institution and promotes its content.

The TIB|AV-Portal benefits from the use of LOD. LOD enables the multilingual annotation of the content, the semantic search, and will enable the retrieval of new related content thanks to the interlinking.

Libraries such as the German National Library, The European Library or Europeana are currently working with LOD in aggregation scenarios and providing the data to third parties. Big broadcasting companies such as the BBC also use Linked Data to annotate their content. With the TIB|AV-Portal, the TIB aims to do the same and weave into a linked ecosystem.

There is still a long way to go, particularly in the use of LOD for non-textual materials. However, there are already successful use cases¹³ generated after the Library Data Incubator.

⁸ <http://lov.okfn.org/>

⁹ <http://www.theeuropeanlibrary.org/>

¹⁰ http://www.filmarchives-online.eu/partners/iwf-knowledge-and-media-1/view?set_language=en

¹¹ www.w3.org/TR/html-rdfa/

¹² <http://lod-cloud.net/>

¹³ <http://www.w3.org/2005/Incubator/lld/wiki/UseCases>

References

1. Berners-Lee, Tim: *Linked Data-Design Issues*, 2006.
2. Heath, Tom, Bizer, Christian: *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool, 2011.
3. Löwgren, Jonas: **"The ground was shaking as the vehicle walked pasted me." The need for video in scientific communication.** *Interactions*, vol. XVIII n.1, pp. 22–25, 2011. DOI: 10.1145/1897239.1897246
4. Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma: Open data y Linked open data: su impacto en el área de bibliotecas y documentación. *El profesional de la información*, 2011, vol. 20, n. 2, pp. 164-172. DOI: 10.3145/epi.
5. Pohl, Adrian, Danowski, Patrick: *Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick*, 2013.
6. Ríos-Hilario, Ana; Martín-Campo, Diego; Ferreras-Fernández, Tránsito: Linked data y linked open data: su implantación en una biblioteca digital. *El caso de Europeana*. *El profesional de la información*, 2012, vol. 21, n. 3, pp. 292-297.
7. Sack, Harald, Waitelonis, Jörg: **Linked Data als Grundlage der semantischen Videosuche mit Yovisto**, in T. Pellegrini, H. Sack, and S. Auer (Hrsg.): *Linked Enterprise Data – Management und Bewirtschaftung vernetzter Unternehmensdaten mit Semantic Web Technologien*, Berlin: Springer Berlin, 2014, pp. 263-287, ISBN: 978-3-642-30273-2.
8. Saorín, Tomás. Cómo linked open data impactará en las bibliotecas a través de la innovación abierta. *Anuario ThinkEPI*, 2012, vol. 6, pp. 288-292.
9. Strobel, Sven, Marín-Arraiza, Paloma: Metadata for Scientific Audiovisual Media: Current Practices and Perspectives of the TIB|AV-Portal. E. Garoufallou et al. (Eds.): *MTSR 2015, CCIS 544*, pp. 159–170, 2015. DOI: 10.1007/978-3-319-24129-6_14
10. Strobel, Sven: Englischsprachige Erweiterung des TIB|AV-Portals. Ein GND/ DBpedia-Mapping zur Gewinnung eines englischen Begriffssystems. In: *obib. Das offene Bibliotheksjournal*, vol. 1, pp. 197-204, 2014. DOI: 10.5282/o-bib/2014H1S197-204.
11. Waitelonis, Jörg, Sack, Harald: Towards Exploratory Video Search Using Linked Data. In: *Multimedia Tools and Applications* 53, pp. 1-28, 2011. DOI: 10.1007/s11042-011-0733-1.
12. Zuiderwijk, Anneke, Jeffrey, Keith, Janssen, Marijn: The Potential of Metadata for Linked Open Data and its Value for Users and Publishers. In: *JeDEM*, 2012, vol. 4, n. 2, pp. 222-244.

Extracting value from grey literature: processes and technologies for aggregating and analyzing the hidden “big data” treasure of organizations^{*}

Gabriele Motta, Roberto Puccinelli, Lisa Reggiani, and Massimiliano Saccone (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.

^{*} First published in the GL17 Conference Proceedings, February 2016.

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 target¹: 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": *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.

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].

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

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

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 planning³;
- 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.

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].

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 multidisciplinary;
- 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.

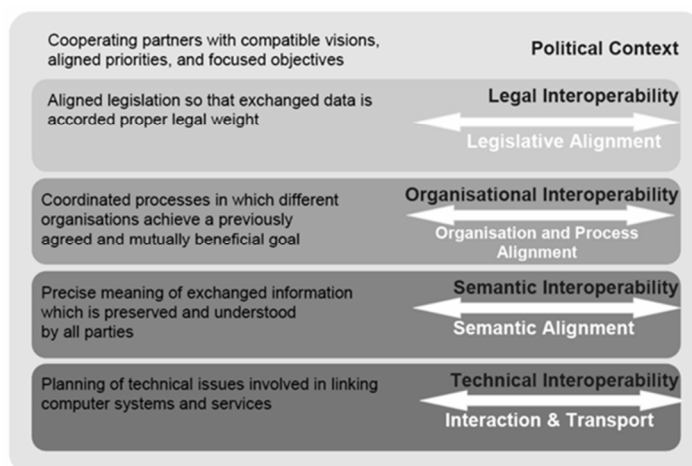


Figure 1 - European Interoperability Framework (EIF) [13]

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.

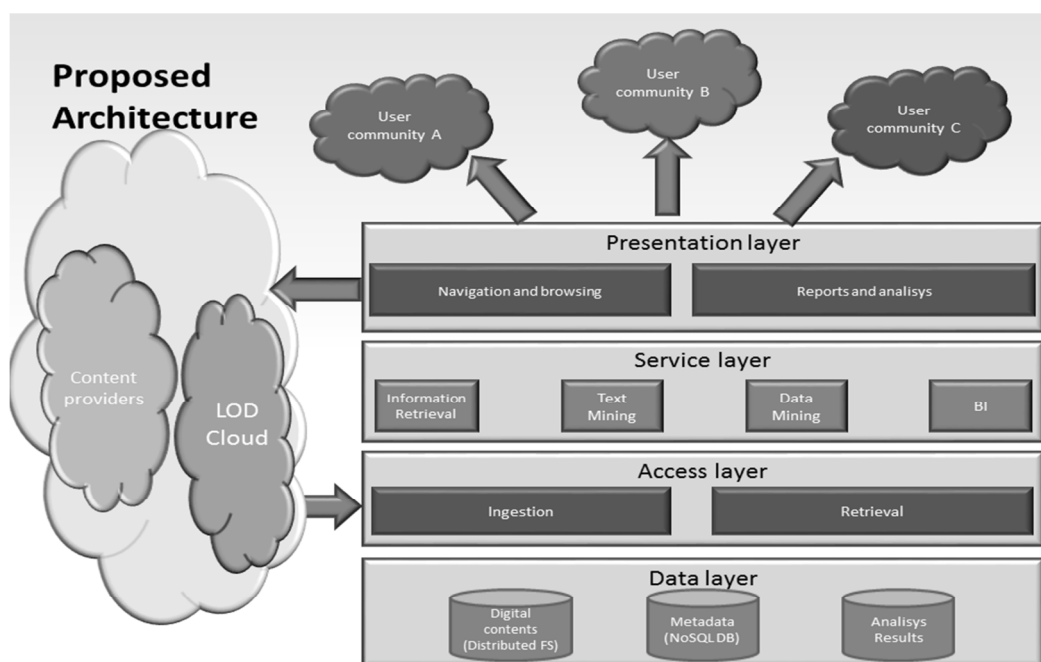


Figure 2 - Proposed architecture (logical view)

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.

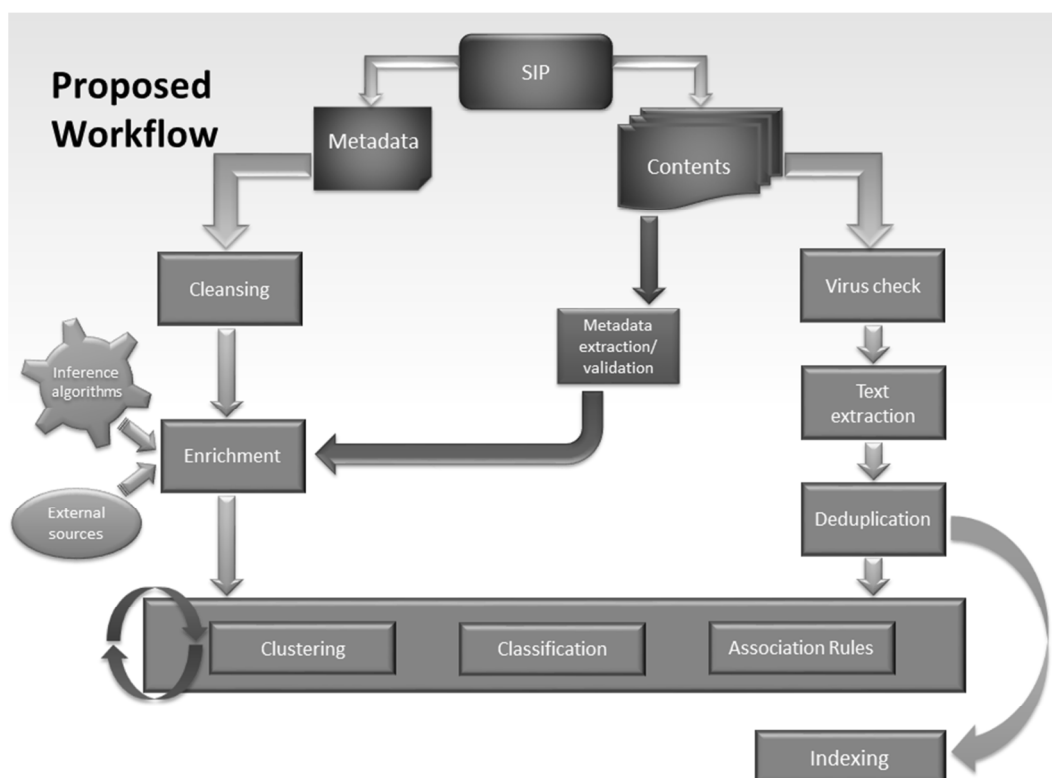


Figure 3 - Proposed core workflow

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 Kacfa 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 (Netherlands)

Stefania Biagioni and Carlo Carlesi (Italy)

Christiane Stock (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 2015¹ 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.

^{*} First published in the GL17 Conference Proceedings, February 2016.

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:

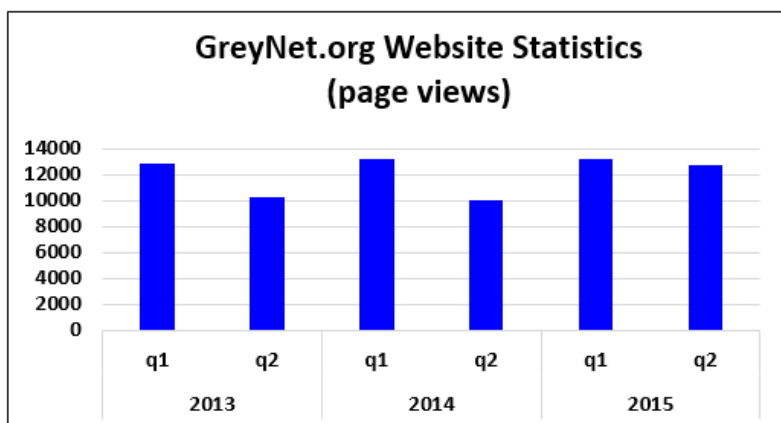
- 2003 – Relaunch GreyNet.org²
- 2003 – Relaunch GL-Conference Series³
- 2004 – GreyNet Annual Award⁴
- 2005 – The Grey Journal, TGJ⁵
- 2007 – Conference Preprints in OpenSIGLE Repository^{6, 7} (later OpenGrey⁸)
- 2009 – GreyNet Workshop Series^{9,10}
- 2010 – GreyNet Social Media¹¹
- 2012 – GreyNet Datasets^{12, 13} in DANS Data Archive¹⁴
- 2013 – GreyGuide^{15, 16} Repository (and Web Access Portal)¹⁷

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

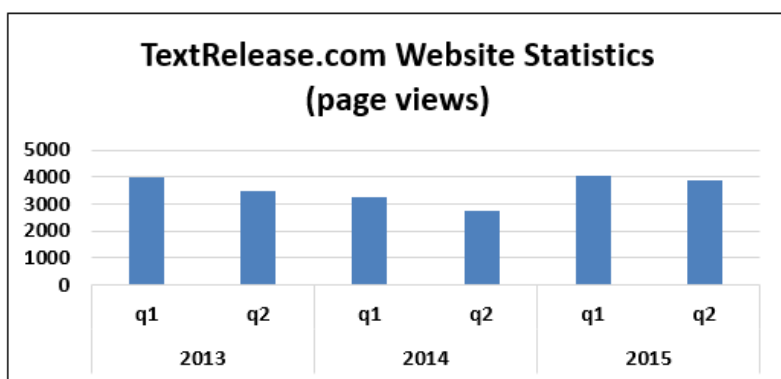
Available data and statistics upon which this paper is based are drawn from a number of sources, namely • GreyNet's In-house Excel and Outlook files • Conference Evaluation Forms • Web Stats (Network Solutions) • Licensed Journal Stats (EBSCO) • Service Provider Stats (OpenGrey, DANS, and GreyGuide)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



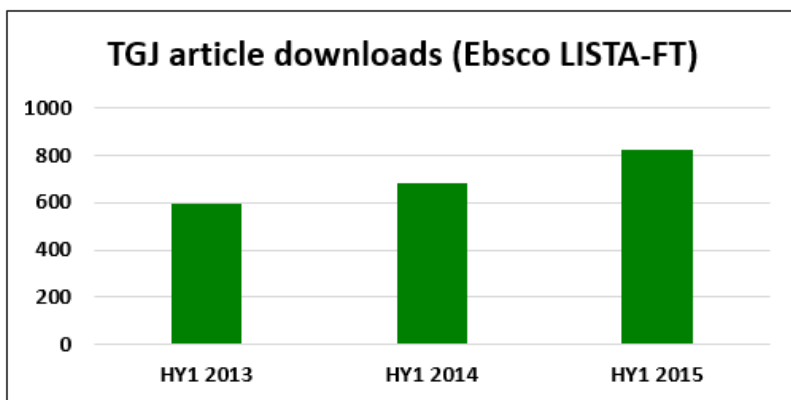
HY1 2013	HY1 2014	HY1 2015
23115	23232	25921

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



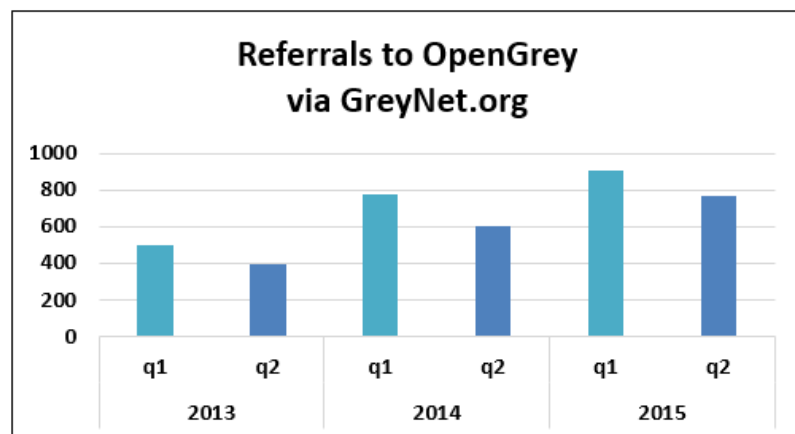
HY1 2013	HY1 2014	HY1 2015
7484	6017	7972

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

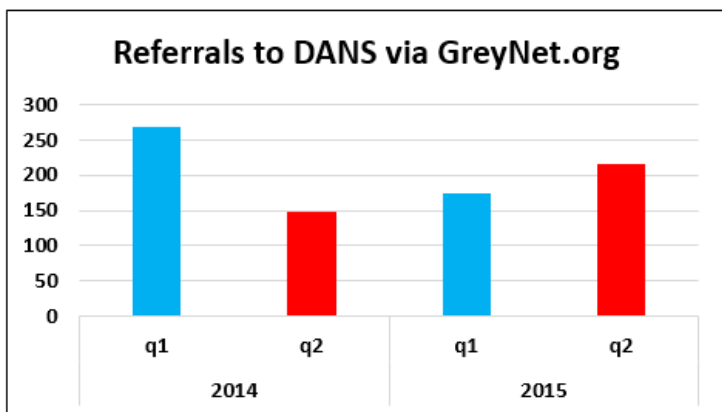


Table 5: Quarterly and half-year totals via GreyNet's website to the DANS Data Archive

GreyNet Referrals to DANS	
HY1 2014	HY1 2015
417	390

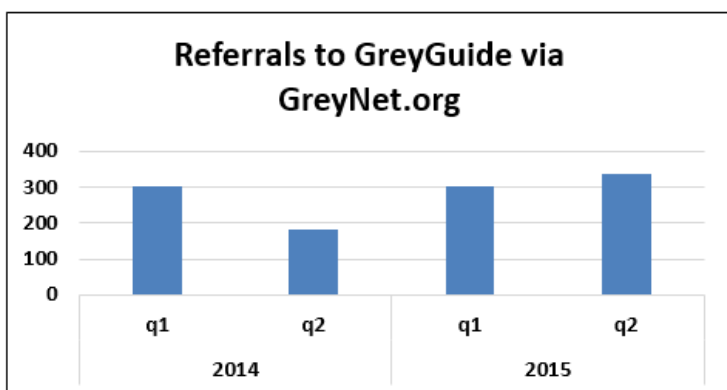


Table 6: Quarterly and half-year totals via GreyNet's website to the GreyGuide Repository

GreyNet Referrals to GreyGuide	
HY1 2014	HY1 2015
487	639

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 participated?

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?

Answer Choices	Responses	
▼ 1-2	29.69%	19
▼ 3-4	7.81%	5
▼ 5 or more	12.50%	8
▼ None	50.00%	32
Total		64

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 communities¹⁸ as well.

References

- ¹ http://www.greynet.org/images/GreyNet_Business_Report_2015.pdf
- ² <http://www.greynet.org/>
- ³ <http://www.textrelease.com/>
- ⁴ <http://www.greynet.org/greynetmembership/annualaward.html>
- ⁵ <http://www.greynet.org/thegreyjournal.html>
- ⁶ OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal / D.J. Farace, J. Frantzen, J. Schopf, 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
- ⁷ 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. Schopf - In: The Grey Journal: An international journal on grey literature. - Amsterdam: TextRelease, Vol. 6, No 1, Spring 2010, pp. 41-48. - ISSN 1574-1796
- ⁸ <http://www.opengrey.eu/search/request?q=greynet>
- ⁹ <http://www.greynet.org/researchandeducation/workshops.html>
- ¹⁰ <http://www.greynet.org/greyforumseries.html>
- ¹¹ <http://www.greynet.org/home/socialmedia.html>
- ¹² 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, Laurens Sesink, and Debbie Rabina. – In: GL13 Conference Proceedings, – Amsterdam: TextRelease, 2012. – pp. 143-150. - ISBN 978-90-77484-17-3
- ¹³ 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
- ¹⁴ <https://easy.dans.knaw.nl/ui/?wicket:bookmarkablePage=:nl.knaw.dans.easy.web.search.pages.PublicSearchResultPage&q=greynet>
- ¹⁵ 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
- ¹⁶ 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
- ¹⁷ <http://greyguide.isti.cnr.it/>
- ¹⁸ Grey communities: An empirical study on databases and repositories / Helene Prost and Joachim Schopf. – In: The Grey Journal, Volume 10, Issue 1, 2014, pp. 38-49. – ISSN 1574-1796

Geologic Field Trip Guidebooks: Shades of Grey^{*}

Lura E. Joseph (United States)

Abstract

Geologic field trip guidebooks contain valuable information useful for both professional geologists and the general public, yet they exhibit various “shades of grey” when attempting to identify and access copies. They are produced by various groups, including governments, professional societies and associations, commercial publishers, and others. Each group presents different levels of difficulty in identifying and accessing related guidebooks.

Research indicates that there are significant indexing gaps for guidebooks from trips held in North America, and the problem is even more profound for guidebooks from international conferences. Search strategies include using multiple resources and search terms.

Descriptors: guidebooks; geologic field trips; geological conferences; geological excursions; geological literature; grey literature

Geologic Field Trips

Most geological conferences consist of indoor presentations of current research, but related field trips are also offered for participants. For many geologists, the earth is their laboratory, and they need to occasionally get outside to see geology in the real world. Large conferences usually offer multiple field trips. There are also “field conferences”, the focus being on the field trip rather than indoor presentations. Field trips and field conferences can last from half a day to a week or longer. There are also “field seminars”, offered as continuing education by industry, or as courses by university geology departments. Many geology departments also require enrolment in a “field school”, often lasting more than a month, during which students learn to apply what they have learned in class to the real world.

Geologic Field Trip Guidebooks

Except for field schools, most geologic field trip leaders provide attendees with a guidebook. Students often create the guidebook for their field seminars under the direction of their instructor. Field trip guidebooks can consist of illustrations copied from other publications (often without permission) and stapled together, handouts at each stop, or more formal, bound publications. Some guidebooks are commercially published for the general public, rather than as part of a specific meeting. A road log is generally recommended so that the trip can be replicated. However, GPS coordinates, river miles, or other methods of trip replication are also acceptable. The Appendix of this paper is a bibliography of information about geologic field trip guidebooks.

Value

Geologic field trip guidebooks are valuable information resources for a number of reasons. They provide geologic information for a particular geographic area. They usually contain general information about the regional geologic setting, as well as specific information about each stop on the trip. Some focus on a particular geological topic such as mining geology, environmental geology, structural geology, or glacial geology. Most include maps, photographs, and other illustrations; many include tables and graphs.

When putting together a field trip, consulting guidebooks from previous trips can reduce the amount of work for the trip leader. Over time, many of the geologic features have been, or will be mined out of existence, covered with concrete, asphalt, or buildings, or made inaccessible by private land owners or commercial enterprises. Field guides are valuable as records of endangered geomorphic features such as moraines, dunes, beaches, and channel fill. A particular guidebook may be the only place that certain information or data are ever published, an example being Kieffer et al (1989). Some groups meet annually to present current field work that may never get published in the detail discussed during the field trip, and the guidebook may serve as the only record of these discussions. Guidebooks are also part of the historical record of the various geological societies.

Geologic field trip guidebooks are useful not only for researchers and scholars, but also for the general public. Many guidebooks contain technical information, but most also contain information that can be understood by the educated nongeologist, and some series are produced specifically for the nonspecialist. Guidebooks are useful for individuals and families traveling to national parks such as the

^{*} This article is expanded and updated from a presentation at the 11th International Conference on Grey Literature, 2009 (Geologic Field Trip Guidebooks; A Project to Identify Indexing Gaps, <http://greyguide.isti.cnr.it/index.php/46-gl11/gl11-slide-share/279-gl11-joseph>).

Grand Canyon, Mt. Rainier, or Smoky Mountains; through the various states; and to international areas having interesting geology. Guidebooks can help satisfy travelers' curiosity about landforms and the rocks exposed in road cuts as they drive along. They can be used to interest children and youngsters about science and ecology.

Sources

There are a number of different producers of geologic field trip guidebooks, and the ease of identifying and locating a copy depends largely on what group or individual produces the guidebook. American Geosciences Institute maintains a list of many of the guidebook series (<http://guide.georef.org/dbtw-wpd/help/series.htm>). Groups that produce geologic field trip guidebooks can be divided into the following: federal and state government surveys, professional societies and associations, academies of science, museums, commercial publishers, university and college departments, and other (including self-publications).

Government

Many federal and state government geological surveys produce field trip guidebooks. Some surveys produce one or more guidebook series, for example, one series for professional geologists, and another for the general public. Other guidebooks are hidden within other survey series such as bulletins, open file reports, and educational booklets. Many, but not all of the survey publications can be identified by searching various indexes; this will be covered later in this article. Most surveys also have lists of their publications on their websites, however it is tedious and time consuming to go to the various survey websites in order to identify guidebooks for a particular topic or area. Some surveys provide search engines, while others only maintain lists. Increasingly, surveys are converting publications to digital format and providing links to full text, although some are currently holding back since their print publications are a source of revenue.

Professional societies and associations

Another significant group of guidebooks are those produced by the various geological professional societies and associations. This group also contributes the widest variety of "shades of grey" when it comes to identifying and locating copies of guidebooks. Some of these groups produce high-quality, bound guidebooks, with consistently formatted covers and title pages which make it much easier for librarians to catalog the material, contributes to consistent indexing, and therefore makes them easier to identify and locate. An example of this type of group is the Institute on Lake Superior Geology. Some of the groups also produce a number of extra copies for sale, however the ease of discovering whether copies are for sale, and how to go about purchasing them, varies widely.

At the other end of the scale, some geological societies are extremely informal. They do not have officers or formal membership. As an example, the Friends of the Pleistocene consist of geographical "cells" that meet annually as field conferences to present cutting edge research in the field. The information about the trips is communicated informally by word of mouth, email, and blurbs in other society newsletters. Once an individual attends a trip, their name is on an email list, and for a few subsequent years they will probably receive notice of the next conference; if an individual fails to attend for a few years, their name will likely be dropped from the notification list. Sometimes no guidebook is produced. There may or may not be material handed out at various stops. Often, if a guidebook is produced, it is a compilation of material from previous publications, and often no permission was obtained to reproduce the information. For this reason, leaders may be reluctant to have the guidebook distributed beyond the trip participants or for the guidebook to be added to a library, or for it to be indexed. This certainly contributes to "a darker shade of grey literature". Authors who are members of government surveys are sometimes reluctant to have their guidebooks distributed beyond participants because of requirements to have work vetted by the agency before publication. In this case, the guidebook will likely contain some sort of disclaimer.

Many other societies and associations fall between these two extremes. Practices pertaining to field trip guidebooks may vary widely between groups, and also within groups. Some society guidebooks are occasionally published by government surveys since guidebook authors may be associated with both groups. This sometimes adds to confusion. There may appear to be a gap in a society guidebook series when, for a particular year, it was published by a survey or co-published by two different societies. Also, two versions of the same guidebook may be "published" – one handed out at the trip with a society series title and number, and one later published by a survey with a survey series title and number. These versions may have identical content, or the content may be revised. To further add to the confusion, a guidebook published by a survey for a society field trip may end up with the same title having two

different volume numbers, which also means one volume number will have two different titles in library catalogs if both volumes end up in libraries. This may happen when a trip leader plans to have the guidebook formally published as part of a survey series, but in the interim between the trip and the survey publication process another publication receives the intended volume number.

Some societies produce a guidebook for the field trip participants, and then modify the information to be included in a bound volume that incorporates all of the trips offered for the conference or the year. The information in the more formal, combined volume is often in less detail than the material that was handed out to participants, but on the other hand, may be more refined. As an example, Geological Society of America publishes a guidebook series for the field trips offered during their conferences. Having two different versions of the guidebook, both of which may end up in libraries, can be confusing and contributes to difficulty in identifying and locating guidebooks, however, it is commendable when societies make the effort to ensure that some version of the guidebook is readily available.

Commercial publishers

Some guidebook series are produced by commercial publishers. These are geared toward non-specialists, although geologists may find them useful. The quality of information in the individual titles varies by author. Examples include two series produced by Mountain Press: the Roadside Geology series, and the Geology Underfoot series. Commercially produced guidebooks can usually be readily identified and obtained.

Other sources

By far, most of the geologic field trip guidebooks are produced by the three groups discussed above. Other groups producing guidebooks include academies of science, museums, and university and college departments. For the most part, these guidebooks are produced sporadically, and therefore can be easily missed by library acquisitions. Also, the guidebooks are included in series with other types of materials, and somewhat “hidden” in much the same way as guidebooks that are included in survey series such as bulletins and open file reports, unless the title or cataloging and indexing clearly indicate that a piece is a geologic field trip guidebook. Guidebooks produced by university or college departments have special considerations. Most are either produced by faculty for department-led trips, or by students for field courses. These are less likely to find their way into library collections, and if they do, there will probably be only one copy in existence at the local institution library. Also, the quality of the student-authored guidebooks may vary considerably.

Another factor contributing to “shades of grey” relative to guidebook literature

Some of the various difficulties related to identifying and obtaining geologic field trip guidebook have been related above. Another consideration is whether guidebooks are included in library collections. If they are not in libraries, they will not be in library catalogs, and also will be less likely to appear in indexes, decreasing the likelihood that they will be identified, and contributing to “a darker shade of grey” literature. Many university and public libraries are governed by accounting policies that make it difficult to purchase material from individuals or small publishers such as societies. State accounting systems usually require the tax numbers of “vendors”. Some small societies do not have a tax number, and individual society officers are understandably reluctant to provide their social security numbers as proxies. Also, many university accounting systems require “vendors” to fill out long forms in order to be listed as an accepted vendor; this is especially true for “foreign” vendors. Some university libraries are able to use a credit card for small orders in order to avoid some of the red-tape; others don’t have that ability. Due to these factors, some university libraries simply give up getting these materials, reducing the number of guidebooks available to the public, and contributing to a “darker shade of grey” literature.

Research

Members of the Geoscience Information Society have made a concerted effort to add geologic field trip guidebooks to their university library collections. Early in the society’s history, members began publishing Union List of Geologic Field Trip Guidebooks of North America. The 6th edition was converted to an online resource (referred to as Geologic Guidebooks of North America, or GGNAD or Guidebook Database) that is continually updated as new guidebook titles are reported. Also, titles related to areas outside of North America began being added to the database as they are reported. However, increasingly, university geology libraries are being closed and materials are being consolidated into other libraries, and experienced geology librarians are retiring. For those and other reasons, few

guidebook titles are being reported. This resulted in a considerable gap in the GGNAD. Previously, GeoRef, the major index and abstracting service for the subject of geology, was not consistently indexing field trip guidebooks, although this has recently changed. There is now a project under way to merge the GGNAD into GeoRef.

North American Guidebooks

In 2008, this author completed a six-month sabbatical project (Joseph, 2009) to determine mutual and unique gaps in three indexes used to identify and locate geologic field trip guidebooks, namely the GGNAD (Guidebooks Database), GeoRef, and WorldCat. The research was focused on guidebooks produced by societies and associations. A series list spreadsheet was created for the societies and associations group, and information was added from the Guidebooks Database, including number of field trips, number of years, and last year in the database. WorldCat was consulted as a further indication as to whether a series was continuing. An arbitrary cut-off date of 1997 was selected, therefore guidebooks series having a guidebook produced in the years 1998 to 2008 were considered to be active. Of the 632 Associations/Societies series identified, only 199 series were considered "active". After eliminating some for various reasons, 105 guidebook series were used for the study.

Spreadsheets were created for each of the series, including guidebook title and year. GeoRef was searched for the series to determine 1) whether the individual titles in the Guidebooks Database (GGNAD) were also in GeoRef, and 2) to identify guidebooks missing from the Guidebooks Database. WorldCat was searched to identify guidebooks missing from the Guidebooks Database and GeoRef. When time permitted, other information sources were consulted, including society lists in guidebooks, web pages of societies, and newsletters, as well as e-mailing field trip leaders and other individuals directly.

There were 5982 total known trips/guidebooks identified for the 105 series. Of the known trips/guidebooks, 1432 (24%) were only in the Guidebooks Database, 823 (14%) were only in GeoRef, 2509 (42%) were in both databases, and 1218 (20%) were in neither database. The 1218 items in the "Neither" category include guidebook titles only in WorldCat, and titles missing from all three databases but reported by librarians and added to a waiting list, or discovered by other means such as lists on society web pages. Searching the Guidebook Database would have identified approximately two thirds of the known trips/guidebooks, and searching GeoRef would have identified somewhat over half. However, the two databases also vary according to time period: The Guidebook Database is more complete for older titles, and GeoRef is more complete for newer titles.

Most significantly, 20% of the known guidebooks/trips would have been missed even if both GeoRef and the Guidebooks Database had been searched. Further, even after searching all three databases and other resources, there could be 1616 or more guidebooks that are yet to be identified. This is based on the finding that there were 1616 "years" among the 105 series for which no trips or guidebooks were identified.

International Conference Guidebooks

The author recently extended the previous project to examine geologic guidebooks for trips related to international geological conferences (Joseph, 2015). As far as determined, no one has ever attempted to systematically study guidebook series for geologic trips outside of North America on a worldwide basis.

Methods refined during the past sabbatical were used in this project. A list of international geoscience conference series was compiled using a variety of resources including lists maintained by umbrella organizations such as American Geoscience Institute. In addition, quarterly issues for the past five years (2010-2014) of the geological magazine *Episodes*, *Journal of International Geoscience* were examined to identify conference series, meetings, and guidebook titles. A spreadsheet was created for each series, and data were recorded on a summary spreadsheet. After individual meetings were identified, various society and meeting websites were examined to determine whether there were associated field trips, and whether there was any mention of associated guidebooks.

During the study, it became apparent that many of the websites no longer existed or else had little or no information about the associated field trips. In some cases limited information was found on other society web pages that included calendars of events of interest to their members. GeoRef and WorldCat indexes were examined using a variety of search strategies to determine indexing gaps.

A total of 230 international geoscience conference series were identified. There were 193 (83.9%) conference series that had at least one meeting in the five years studied (2010-2014). The total number of international meetings identified during that five year period was 414.

Of the 414 known meetings, 204 (49.3%) had at least one identified field trip. In the author's experience, having attended numerous professional geology conferences over more than a 40 year span

of time, most geology/geoscience meetings have at least one associated field trip. Assuming all 414 identified meetings did have at least one field trip, about 210 guidebooks may exist that could not be identified (50.7% of 414). However, there are often multiple trips per meeting, so the actual total number of unidentified field trips could be much higher.

For the 414 identified meetings, the total number of known field trips discovered was 701. Only 146 guidebooks were located for the 701 identified field trips (20.8%). That is a remarkably low percentage. In the author's experience, most field trips have at least some sort of guidebook, therefore as many as 555 guidebooks from this sample of field trips might be unidentified. Added to the 210 that might be unidentified due to unidentified field trips, the total could be at least 765 unidentified guidebooks. Therefore, 84% of guidebooks may be unidentified from this 5-year sample of international geoscience conferences.

Of the total 146 guidebooks identified, the number of guidebooks that can be identified in WorldCat is 101 (69.18%), and the number that can be identified in GeoRef is 56 (38.36%). Multiple search strategies were used and included the society name, the destination of the field trip, and possible wording of guidebook titles. It is possible that additional titles are contained in the indexes, but without more information regarding the titles, it is unlikely that additional guidebooks could be identified.

This research indicates that geoscience fieldtrip guidebooks from international conferences, though valuable, is extremely endangered literature. While the guidebooks from North America have had the attention of the geology librarians, it appears there have been no stewards of the international segment of this literature. Information from earlier meetings is probably even more difficult to find, since the persistence of these web pages likely decreases with time.

Many societies are not maintaining or archiving past meeting websites, making it very difficult to identify past meetings and meeting details. Even when old meeting websites are archived, they may lack information regarding fieldtrips. Even when the possibility of field trips is mentioned, it is often impossible to determine if a trip was actually held, or what the title of the trip was, and whether there was a guidebook and what the title of the guidebook might have been. Without that information, it is less likely that a guidebook will be identified in the indexes and library catalogs.

Suggestions for identifying and locating guidebooks

There are several tools available for seeking field trip guidebooks (Joseph, 2006). No single information source is complete. Three of the best resources to search are GeoRef, WorldCat, and the Geologic Guidebooks of North America (GGNAD). The GGNAD is currently being merged into GeoRef. Each database may require modification of the search strategy. As with any database, multiple search strategies may be required. Search terms may include both the singular and plural of: guidebook OR guide book OR guide; field trip OR fieldtrip OR excursion. (Excursion is often used in place of field trip for non-North American trips, but may result in false hits because it is also used as a graphical term). The terms geology OR geologic OR geological OR geoscience may be needed to refine searches in multidisciplinary databases such as WorldCat. Depending on the desired outcome, it may be necessary to search using different geographical terms and subdiscipline terms, multiple series and guidebook titles, leaders, authors, etc.

In addition to the above suggestions, other resources include the web pages and publications of geological societies and surveys, and web browser searches. Some guidebook series include a list of previous trips. More detailed suggestions for searching can be found in an article by Joseph (2006).

Conclusions

Geologic field trip guidebooks contain valuable information for researchers and scholars, as well as the general public. Whether doing a broad search, a focused exhaustive search, or seeking a specific individual title, geologic field trip guidebooks cover the range of "shades of grey" literature. Some commercial series and titles are easily identified and accessed, while other types are nearly invisible. Research indicates that international conference guidebooks may be in even more jeopardy than those for North American trips. Science librarians, as well as international geological societies and umbrella organizations need to be involved in preserving and indexing this important body of information.

References

- Joseph, Lura E. 2006. Where are geologic field trip guidebooks when you need them? *Geoscience Information Society Proceedings* 37: 23-29.
- Joseph, Lura E. 2009. Geologic field trip guidebooks; Progress on a project to identify indexing gaps. *Geoscience Information Society Proceedings* 40: 11-17.

- Joseph, Lura E. 2015. Guidebooks associated with international geological conference field trips; endangered? Geological Society of America presentation Abstract No: 259576 (paper expected to be published in Geoscience Information Society Proceedings, 2016).
- Kieffer, S.W., Graf, J.B., and Schmidt, J.C. 1989. Chapter 3: Hydraulics and sediment transport of the Colorado River, in Elston, D.P., Billingsly, G.H., and Young, R.A., eds., *Geology of Grand Canyon, Northern Arizona* (with Colorado River Guides): Lees Ferry to Pierce Ferry, Arizona: 28th International Geological Congress, American Geophysical Union, p. T115/315:51.

APPENDIX - Bibliography of Information about Geologic Field Trip Guidebooks

- American Geological Institute/Geosciences Information Society. Current. Geologic Database of North American Guidebooks: <http://www.agiweb.org/georef/online/bgnaintro.html> (Accessed 13 October 2015).
- Bichteler, J. 1991. Geologists and gray literature: access, use, and problems. *Science and Technology Libraries* 11:39-50.
- Butkovich, N.J., Gomez, J., and Baker, V. 1992. Formats of cited references in geological journal literature. *Collection Management* 16 (2): 61-73.
- Corbett, R.G. 1989. Field trip guidebooks need not be gray literature: *Geoscience Information Society Proceedings* 19:113-122.
- Derksen, C.R.M. 1991. Overlooked sources of information: Geologic field trip guidebooks. *Compass* 68:100-103.
- Foote, Jody Bales. 2010. State geological survey libraries: a disparity in resources, services, access, and professionalism. *Science & Technology Libraries* 29 (1-2): 53-68.
- GSIS, Geoscience Information Society Union List of Field Trip Guidebooks Committee, eds. 1996. *Union list of geologic field trip guidebooks of North America*. Alexandria, VA: American Geological Institute.
- GSIS, Geoscience Information Society Guidebook Standards Subcommittee. 2005. Guidelines for authors, editors, and publishers of geologic field trip guidebooks: Geoscience Information Society <<http://www.geoinfo.org/GuidebookGuidelines.pdf>> (Accessed 13 October 2015).
- Haner, B. E. 1990. Guidebook citation patterns in the geologic journal literature: a comparison between 1985 and 1967. *Geoscience Information Society Proceedings* 20, p. 159-169.
- Holoviak, J. C. 1990. Publishers' decisions in the field trip guidebook business. *Geoscience Information Society Proceedings* 20:199-203.
- Johnson, K.S. and D.J. Johnson. 1990. Ingredients of a great geologic field trip. *Geoscience Information Society Proceedings* 20:205-208.
- Joseph, L.E. 2002. Help libraries preserve field trip guidebooks. *The Quaternary Times* 32 (2):12.
- Joseph, L.E. 2004. The web-based academic field trip bibliography: a multi-use library tool. *Issues in Science and Technology Librarianship summer*, article 6 <<http://www.istl.org/>> (Accessed 13 October 2015).
- Joseph, Lura E. 2006. Where are geologic field trip guidebooks when you need them? *Geoscience Information Society Proceedings* 37: 23-29.
- Joseph, Lura E. 2007. Quaternary research. Chapter 3 in *Literature Search Strategies for Interdisciplinary Research*, edited by Linda Ackerson. Lanham, Md.: Scarecrow Press, 31-45.
- Joseph, Lura E. 2009. Geologic field trip guidebooks: progress on a project to identify indexing gaps. *Geoscience Information Society Proceedings* 40: 11-17.
- Joseph, Lura E. 2015. Guidebooks associated with international geological conference field trips: endangered? Geological Society of America presentation Abstract No: 259576 (paper expected to be published in Geoscience Information Society Proceedings, 2016).
- Litherland, Kate, and Stott, Tim A. 2012. Virtual field sites: Losses and gains in authenticity with semantic technologies. *Technology, Pedagogy and Education* 21 (2): 213-230.
- Melhorn, Wilton N., and Laffoon, Carolyn J. 1987. Perceptions from beyond the stacks: a user view of the geoscience library. *Geoscience Information Society Proceedings* 17: 3-9.
- Sharp, R.P. 1988. Earth science field work: role and status. *Annual Review of Earth and Planetary Science* 16: 1-19.
- Stark, Marilyn M. 1985. Professional publishing as a source of geoscience information. *Geoscience Information Society Proceedings* 16: 159-166.
- Stott, Tim. 2010. Design, development and preliminary student evaluation of virtual field guides as aids to teaching and learning in the earth sciences. *Geophysical Research Abstracts* 12: 1107. (EGU2010-1107-1, EGU General Assembly 2010).
- Stott, Tim, Nuttall, Anne-Marie, and McCloskey, Jim. 2009. Design, development and student evaluation of a virtual Alps field guide: *Planet* 18, 18-21
http://www.researchgate.net/publication/239588372_Design_development_and_student_evaluation_of_a_Virtual_Alps_Field_Guide (Accessed 13 October 2015).
- Stout, D.L. 1990. Preparing a Kenyan field guide: *Geoscience Information Society Proceedings* 20: 209-212.
- Thompson, Thelma B. 2010. New England Intercollegiate Geological Conference guidebooks: Part one: evolution. *Granite State Geologist* 71 (Fall-Winter 2010): 2-4. http://www.gsnhonline.org/publications/Documents/newsltr_2010-12.pdf (Accessed 13 October 2015).
- Thompson, Thelma B. 2010. The role of field trip guidebooks in field-based geologic research: examining sixty years of NEIGC guidebooks: Geological Society of America Abstracts with Programs, Abstract 16947.
https://gsa.confex.com/gsa/2010NE/finalprogram/abstract_168947.htm (Accessed 13 October 2015).
- Walcott, R. 1990. Guidebook problems from the librarian's point of view: *Geoscience Information Society Proceedings* 20: 185-192.
- Wallace, H.E. 1978. Geological field trip guidebooks: bibliographic problems. *Special Libraries* 69: 291-295.
- Zipp, L.S. 2000. Setting preservation priorities: the geologists' perspective. *Geoscience Information Society Proceedings* 34:55-64.

Grey Literature Sources in Historical Perspective: Content Analysis of Handwritten Notes^{*}

Snježana Ćirković (Serbia)

Abstract

This paper describes historical perspective of grey literature in comparison to the postulates of Serbian enlightenment in the 18th century. The theoretical part of this paper describes comparative analysis between main postulates of grey literature given in *Pisa Declaration on Policy Development for Grey Literature Resources* (Pisa, Italy) on May 16, 2014 in comparison to the first manifest of Serbian enlightenment written by one of the greatest Serbian enlightener Zaharija Orfelin. Orfelin has written this manifest in the preface of the Journal *Slaveno-Serbian Magazin*, founded in Venice (Italy) in 1768. The second part of this paper describes the handwritten notes, as one of the grey literature forms, made by Zaharija Orfelin on the books in his private library. This library presents one of the first private Serbian libraries from the 18th century, it is a rich collection of the most representative books from that period in different languages (German, Italian, French, Russian, Latin and Slaveno-Serbian) and represents invaluable cultural heritage. Using the method of content analysis, this paper describes these handwritten notes, and their role as grey literature source in reconstructing the biography of Zaharija Orfelin compensating for lack of other primary biographical and historical printed sources.

Introduction

The 18th century is the century of migrations and great exodus of the Serbian people and re-establishment of Serbian national identity - the century in which modern Serbian civic culture was born. One of individuals, who contributed to this creation, was Zaharije Orfelin.

Mr Orfelin was a unique personality in his own way, both in terms of erudition and the legacy he left in Serbian art, specifically in the fields of engraving, calligraphy, poetry and education. "His contemporaries glorified him as the most talented and learned among Serbian artists in the early 18th century. He was the first who reached the highest level of engravers' skills and became respected in Vienna, and the sole artist who established himself as a serious historian and wrote a monumental work on Peter the Great" (Todorović, 2006).

Orfelin was also a teacher, owner of the largest library of his time in the Archbishopric and the founder of the first publishing house. He was recognized and respected member of Jacob's Schmutzers Imperial engraving Academy and he received imperial rewards and grants for his calligraphic work. Orfelin's life was filled with unusual diversity and strong contrasts due to both private and professional circumstances. He frequently changed places, countries and posts.

He was an erudite and the list of his occupations is very long. Let me highlight some of them. Orfelin was first Serbian poet who published his poems; he established the first Serbian and South Slavic Journal *Slaveno-Serbian Magazin*, in Venice in 1768; compiled the first Serbian recommended bibliography; wrote the first history of Peter the Great in a Slavic language, and all this for common good, which was the main postulate of his entire work.

Zaharija Orfelin – A Short Biography

Zaharija Orfelin was born in the city of Vukovar, Austrian Monarchy, in 1726. First twenty years of his life are still unknown to the Orfelin's biographers. What we know is that he was a teacher in a Slavic School in Novi Sad, and in 1757 he moved to Sremski Karlovci, where he became the secretary to the Serbian Metropolitan Pavle Nenadovic. This small town in today's Vojvodina was Serbia's most important political and cultural centre at the time. These five years which Orfelin spent in Sremski Karlovci were one of his most productive periods. Together with the Metropolitan Pavle Nenadovic Orfelin founded the *Copper Publishing House*, where he published his first poems, translated books, and created graphics and engravings, inspired by the work of his contemporaries.

Some of Orfelin's biographers assume that his inquisitive mind took him to Venice, where he had an opportunity to publish more books and get in touch with the contemporary West-European thought and ideas of enlightenment.

Other Orfelin's biographers assume that he had to move from Sremski Karlovci, because of political pressure he was exposed to. Namely, in 1761 he published a poem *The Lament of Serbia*. He first published it anonymously in Venice in 1761. In this poem he criticized Austrian establishment and the high ecclesiastical circles in the Archbishopric, while also emphasizing the difficult position and unjust

^{*} First published in the GL17 Conference Proceedings, February 2016.

treatment of the Serbs under the Imperial protection. Despite the fact that it was unsigned, the Orfelin's authorship was soon discovered, which gave him a status of *persona non grata* and had to leave Sremski Karlovci.

In 1764, Orfelin came to Dimitrios Theodosios's publishing house in Venice, where he spent the next six years before moving back to Karlovci. Theodosios, who began to print Serbian books in Venice in 1758, needed Orfelin to edit and proofread the abundant material being brought to him. Dimitrios Theodosios's publishing house remained the only place for publishing of Serbian literature and printed books in Russo-Slavonic until the emergence of Kurtzbek's publishing house in Vienna in 1770. The Orfelin's contributions as a redactor and editor in this publishing house were huge. These six years which Orfelin spent working together with Dimitrios Theodosios are described as a "golden age" of this publishing house. (Ćurčić, 2002)

Slaveno-Serbian Magazin

Slaveno-Serbian Magazin is the first Serbian and Slavic journal published in Venice in 1768. Zaharija Orfelin published this journal anonymously. A role model for this journal was Russian journal *Ežeměsečnaja Sočineniâ*.

Slaveno-Serbian Magazin contained short novels, poems, numerous articles from different scientific disciplines and first Serbian recommended bibliography. The most important part of this journal is the preface, written on 96 pages, and described as the *Manifest of Serbian Enlightenment*. That was the first time that ideas of rationalism, characteristic for the 18th century, were published in Serbian language.

For Orfelin, the 18th century was a century where education and knowledge cease to be any more the privilege of certain social classes, but rather should become available to everyone, to all social classes. Everything is focused on community, common good and humanization of a man as a being. For rapidly growing need for knowledge sharing, books were not any more the best and the fastest way. Journals became much better instrument and form. They offer a great deal of knowledge from different subjects, consist of short articles and offer a lot of different types of information. The magazine form was Orfelin's attempt to compensate for a lack of public schools and books. We can summarize the main postulates of this preface as following:

- Sharing knowledge among all social classes;
- Common good;
- Spreading knowledge through new printed forms (journals, almanacs, manuals) instead of books, in lack of schools and official education;
- Relief from clerical influences.

Although the *Slaveno-Serbian Magazin* was published two centuries before, the ideas standing in background, the enlightenment ideas about spreading knowledge, or better to say "knowledge for everyone", remained till today as an intellectual legacy in many different ways. One of them, appeared some 250 years later, as the *Pisa Declaration on Policy Development for Grey Literature Resources* published in Italy in May, 2014. In this paper I will make a comparative analysis of these two documents, but let us first give a look in Orfelin's library and handwritten notes founded on the books as a grey literature source.

Private Library of Zaharija Orfelin

This library presents one of the first private Serbian libraries from the 18th century, it is a rich collection of the most representative books from that period in different languages (German, Italian, French, Russian, Latin and Slaveno-Serbian), and represents invaluable cultural heritage. Orfelin's library is in every aspect an important source for the better understanding of his personality. Book collection of one man opens a door to his spirit and character, gives us a way to perceive the height, depth and direction of his education, and his interests - literary, scientific and personal one.

When we analyze the character of the Orfelin's library, we see that he was buying some books to expand his intellectual horizons, some of them to gain a practical knowledge and skills, some of them he was buying to educate others, while some he used as sources for writing *the History of Peter the Great*. (Ćirković, 2013)

In the preface to this book Orfelin says: "Many historians, when they want to write a history book, are happy because they can use state, public or academic libraries. For me who lives in faraway regions was needed to travel to Germany and buy German books or to Russia to buy Russian books, and all that as much as was allowed by my funds."

If we accept Orfelin's statement that he was an autodidact, then this rich library, which contains valuable and significant scientific works from various disciplines (Theology, History, Linguistics,

Medicine, Astronomy, Botanic, Arts, Bibliography, Pedagogy, Geography, Physics, Law, Literature, Stylistics, Numismatics, etc.), could be seen as his attempt to compensate the knowledge which he missed during his youth days.

Important additional values of this library are the handwritten notes written by Orfelin himself on the margins of the books. Every book from Orfelin's library has his signature, in a form of Ex Libris, and confirmation that the book belongs to him.

We can find also numerous notes about book content, date and time when and where Orfelin bought the book, even to whom he borrowed it. On many books in his library we find different types of bibliographies and book lists, which Orfelin used as a reference tools for his literary and scientific work.

Handwritten notes as historical sources

Any remnants of the past can be considered as a historical source. That might be a document, a piece of art, some ephemeral or any other object. These are all sources because they provide us in different ways with information which can add to the sum of our knowledge about the past. However sources only become historical evidence when they are interpreted by the historians and put in the historical perspective and contents. There are three main types of historical sources:

- I. **Primary source** - an original document containing firsthand information about a topic (diaries, interviews, letters, handwritten notes, original works of art, photographs, speeches, works of literature);
- II. **Secondary source** - commentary/discussion about a primary source; interpretation of primary sources (biographies, dissertations, indexes, abstracts, bibliographies, journal articles, monographs);
- III. **Tertiary source** - presents summaries or condensed versions of materials, usually with references back to the primary and/or secondary sources (dictionaries, encyclopedias, handbooks, etc.)

Orfelin's handwritten notes - Content analysis

As previously mentioned, Orfelin's library serves as source for better understanding of his personality and his intellectual profile. Numerous handwritten notes give us an insight into his travels, attitudes, habits, personality, literary and scientific interests. Based on typology of most common handwritten notes on the books in Orfelin's library, we can distinct four types of handwritten notes:

1. **Form of Ex Libris** – Signature on the books - *Z.O., Zacharias Orphelin, Zaccaria Orfelini...*;
2. **Notes about the book content** – *Author lies (лжетъ авторъ), materialist (матеріалистъ),* explanation to some terms in books..;
3. **Notes about date and place of book purchase** - *Das Buch habe ich gekauft um zwanzig Groschen von den Buchhaltern 24 July 1751 in Ofen;*
4. **Different types of title lists, short bibliographies** – *КНИГИ на нѣмецкомъ и славенскомъ языкѣ...*

All these notes are important for Orfelin's biographers and historians. Their importance is reflected in following:

- For reconstructing of Orfelin's biography; in lack of other primary biographical sources;
- For creating an intellectual Profile of Zaharija Orfelin;
- For better understanding his personality, habits, personal interests, etc.

Grey Literature – Overview

The metaphor "the chameleon of information resources" maybe describes the grey literature (GL) the best. As the author further states, "it can constitute virtually anything and be written for and by anyone in almost any format" (Rucinski, 2015: 552). The ephemeral and variable nature of grey publication types, editions, and formats makes it hard to describe and define. Despite that, there are many definitions of grey literature. The most widely accepted one is the *Luxemburg definition*. The newest one is Prague definition from *The Twelfth GL International Conference (Prague, 2010)*, which defines GL as follows:

Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body (Schöpfel, 2011: 17)

Grey literature is commonly described as a field in library and Information science that deals with the production, distribution, and access to multiple document types produced on all levels of government, academics, business, and organizations in electronic and print formats not controlled by commercial publishing, i.e. where publishing is not the primary activity of the producing body. Examples of grey literature include patents, technical reports from government agencies or scientific research groups, working papers from research groups or committees, white papers, and preprints.

According to the Wikipedia "Common grey literature publication types include reports (annual, research, technical, project, etc.), working papers, government documents, and evaluations. Organizations that produce grey literature include government departments and agencies, civil society or non-governmental organizations, academic centers and departments, and private companies and consultants.

Grey literature may be made available to the public, or distributed privately within an organization or group, and often lacks systematic means of distribution and collection. The standard of quality, review and production can also vary considerably. Grey literature is therefore often difficult to discover, access, and evaluate."¹

Pisa Declaration on Policy Development for Grey Literature Resources

The Pisa Declaration on Grey Literature was developed at a forum held in Pisa, Italy in May 2014 organized by *GreyNet*² and the *National Research Council of Italy*. It is currently translated into 20 languages worldwide (English, Armenian, Bulgarian, Croatian, Czech, Dutch, French, German, Greek, Hindi, Hungarian, Italian, Japanese, Korean, Macedonian, Russian, Serbian Cyrillic, Serbian Latin, Spanish, Tagalog, Turkish).

*The Pisa Declaration*³ provides a 15-point roadmap that should serve as a guide for organizations involved with the production, publication, access and use of grey literature well into this 21st century.

Until now, the problem was the lack of cooperation and coordination between and among organizations dealing with grey literature. *The Pisa Declaration* marks an end to *ad hoc* policy and decision making with regard to grey literature resources.

The main points set out in *the Pisa Declaration* can be grouped into five categories:

1. Organizational commitment to open access and the sharing of open data standards;
2. Commitment to research and education, where recognition and reward is associated with quality grey literature, and where attention is given to good practices in the field;
3. Commitment to address and safeguard legal issues inherent to grey literature by exploring the various types of licensing agreements now available and by fostering constructive relations with commercial publishers;
4. Commitment to sustainability linked to a financial prerequisite. Identifying funding and grants for special collections and repositories, commitment to long-term preservation, and investments in new technologies;
5. Firm technical commitment to continued online services and further cross-linking of textual and non-textual content – a commitment that ranges from tackling broken links to facilitating interoperability regardless of the system or portal in which grey literature and its accompanying data are housed. (Farace, 2014)

"It is in this way that the Pisa Declaration can reveal in the strengths and opportunities that grey literature offers, while at the same time exposing the weaknesses and threats facing our community. No longer are we resigned that grey literature is hard to find, but instead how can we best search and access it. No longer hold in question its worth and value, but instead set out the review process it has undergone. And, no longer hesitate as to whether it is published or not, but instead cite and reference grey literature – make it openly public isn't that what published means?" (Farace, 2014: 3)

Comparative Analysis – Slaveno-Serbian Magazin and Pisa Declaration

As I have already highlighted, the four main postulates of *Slaveno-Serbian Magazin* are:

- Sharing knowledge among all social classes;
- Common good;

¹ https://en.wikipedia.org/wiki/Grey_literature

² <http://www.greynet.org/>

³ <http://greyguide.isti.cnr.it/>

- Spreading knowledge through the new printed forms (journals, almanacs, manuals) instead of books, in lack of schools and official education;
- Relief from the clerical influences.

These postulates can be also found in *Pisa Declaration*, in the form of open access, knowledge transfer and innovation:

“In order to realize the benefits of research and information for scholarship, government, civil society, education and the economy, We, the signatories to this declaration, call for increased recognition of grey literature’s role and value by governments, academics and all stakeholders, particularly its importance for open access to research, open science, innovation, evidence-based policy, and knowledge transfer.”⁴

Sharing knowledge in the 18th century was one of the essential things that should be done in order to educate all social classes and bring the knowledge to everyone through different and innovative ways. Open access as one of the postulates given in *Pisa Declaration* points at sharing knowledge between scientists, knowledge for everyone, which take us back to the 18th centuries’ enlightenment idea of common good.

Both of these documents were published in Italy, with the time distance of almost 250 years, but the idea of spreading knowledge, innovative ways and forms of knowledge transfer remain the same till today. Mutual interaction of these postulates is presented in the Figure below.

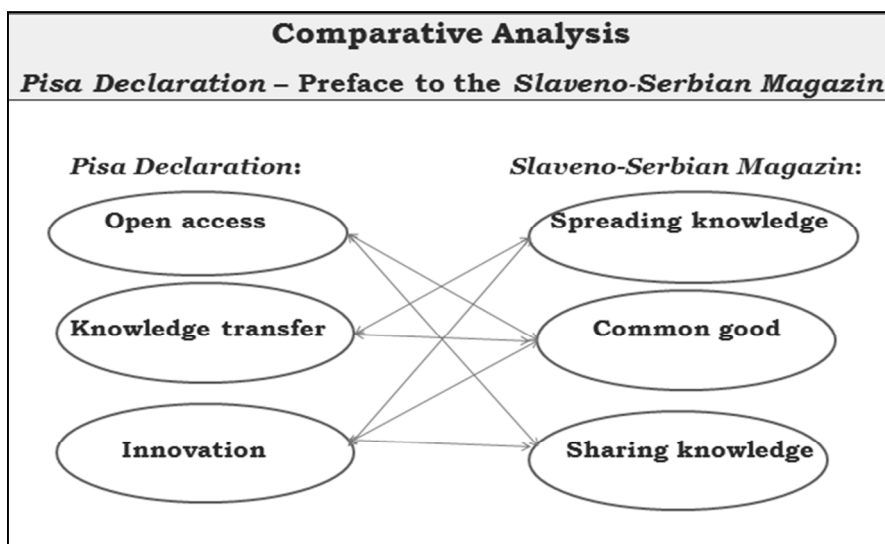


Fig. 1.: Comparative Analysis: *Pisa Declaration* and *Slaveno-Serbian Magazin*

Conclusion

Handwritten notes are important type of grey literature sources and in order to benefit from them, we need to:

1. **Include handwritten notes in corpus with other grey literature sources** (such as: research and technical reports, briefings and reviews, evaluations, working papers, conference papers, theses, and multimedia content);
2. **Develop standards for bibliographical description and implement them in applicable metadata structures;**
3. **Promote use of notes which are today mostly comments.**

Notes and comments as a source of grey literature went at least through four different development phases:

1. In the 18th century there were handwritten notes on the margins of books;
2. In the 20th century we had sticky notes in the books which we used to make highlights and comments, while protecting the books;

⁴ http://www.greynet.org/images/Pisa_Declaration,_May_2014.pdf

3. Today, in 21st century, handwritten notes developed into a new form of comments on eBooks. For example Ebrary, where the software gives us an opportunity to communicate with other users, and where all the notes, although still grey, are openly visible to the broad range of users, and not hidden in library treasures on margins of some old books, waiting to be discovered and tell us some new story. From grey to bright...forms are changing, but message remains.

References

- Blaaij, C. de (2007). The use of grey literature in historical journals and historical research: A bibliometric and qualitative approach. *GL9, Antwerp (Belgium)*. Retrieved from: <http://hdl.handle.net/10068/697875>
- Ćirković, S. (2013). *Biblioteka Zaharija Orfelina*. Beograd : Institut za teološka istraživanja, Pravoslavni bogoslovski fakultet Univerziteta u Beogradu.
- Ćurčić, L. (2002). *Knjiga o Zahariji Orfelinu*. Zagreb : Srpsko kulturno društvo Prosvjeta.
- Farace, D. (2014). The GreyGuide Repository and Web-Access Portal: GreyNet's Response to the Pisa Declaration. *Conference on Grey Literature and Repositories: proceedings 2014: the Value of Grey Literature in Repositories* [online]. Prague: National Library of Technology. Retrieved from: http://repozitar.techlib.cz/record/808/files/idr-808_4.pdf
- Orfelin, Z. (2000). Slaveno-serbski magazin: predgovor. *Ljetopis Srpskog kulturnog društva "Prosvjeta"*. Sv. 5., p. 189-195.
- Rucinski, T. L. (2015). The Elephant in the Room: Toward a Definition of Grey Legal Literature (December 2015). *Law Library Journal*, Vol. 107:4, 2015. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2708884
- Schöpfel, J. (2010). Towards a Prague Definition of Grey Literature. *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010, Dec 2010, Czech Republic*. pp.11-26. Retrieved from: <http://hdl.handle.net/10068/700015>
- Todorović, J. (2006). *An Orthodox Festival Book in the Habsburg Empire – Zaharija Orfelin's Festive Greeting of Mojsije Putnik's in 1757*. London: Ashgate.

ON THE NEWS FRONT

TIB joined GreyNet as an Institutional Member in 2016



TIB is the German National Library of Science and Technology, as well as architecture, chemistry, computer science, mathematics and physics. In its specialist fields it provides national and international researchers and industry practitioners and – as a University Library – students, researchers and teaching staff from all faculties of Leibniz University Hannover as well as interested local citizens with literature and information in both printed and electronic form. The library has outstanding collections of fundamental and highly specialised literature on science and technology. Its remit is to preserve the knowledge recorded and to provide the latest information for research and industry, both now and in the future, irrespective of the time and the place. The aim of TIB is to provide its customers with innovative, efficient and smooth access to data, information and knowledge, and to make large information spaces available to them in a structured, suitable manner. One of its specialist areas is grey literature from its subject areas, which is difficult to procure and cannot be obtained commercially. With its portal, TIB offers a search and order portal for specialist and research information with a focus on science and technology. Knowledge objects such as audio-visual media, 3D models and research data are also integrated in the portal. More than 170 million data set are currently available. TIB conducts applied research and development to optimise the services it offers. In cooperation with research facilities, emphasis is placed on the topics of visual search, the visualisation of data, future internet and semantic web. By allocating DOI names (digital object identifiers) and ensuring they can be detected and accessed at TIB safeguards the quality, long-term availability and referencability of research data. TIB is a member of the Leibniz Association, and became a public-law foundation of the Federal State of Lower Saxony in 2016. <http://www.tib.eu>

GL18 Program Committee Members and Organizations



Danielle Aloia ^{Chair}

The New York Academy
of Medicine, NYAM
United States



Ľubomír Bilský
Slovak Centre of Scientific
and Technical Information
CVTISR, Slovak Republic



Meg Tulloch
FEDLINK
Library of Congress
United States



Marnix van Berchum
Data Archiving and
Networked Services,
DANS-KNAW, Netherlands



Dobrica Savić
Nuclear Information Section
International Atomic
Energy Agency, Austria



Christiane Stock
Institut de l'Information
Scientifique et Technique
CNRS, France



Petra Pejšová
National Library of
Technology
Czech Republic



Stefania Biagioni
Institute of Information
Science and Technologies
CNR, Italy



Deborah Rabina
Pratt Institute,
School of Information
United States



Kiseok Choi
Korea Institute of Science and
Technology Information, KISTI
Korea



Joachim Schöpfel
University of Lille
France



Grey Literature Network Service

Dominic Farace
GreyNet International
Grey Literature Network
Service, Netherlands



Tomas A. Lipinski
University of Wisconsin,
Milwaukee, UWM
United States

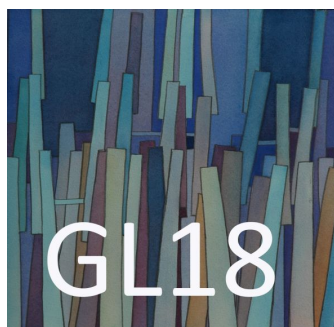
TextRelease

GL18 Program and Conference Bureau

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
www.textrelease.com • conference@textrelease.com
Tel/Fax +31 (0) 20 331 2420

<https://doi.org/10.28069/greyret-2025-000.538-tqj>

Eighteenth International Conference on Grey Literature Leveraging Diversity in Grey Literature



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.

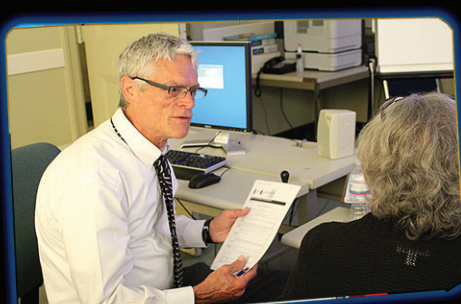
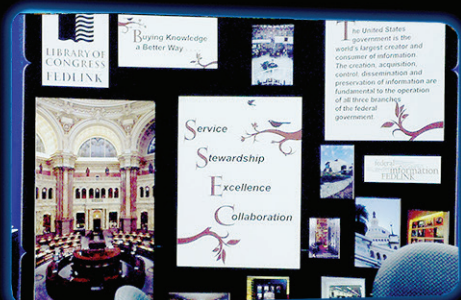
TextRelease

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

FEDLINK

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



Strategic Sourcing



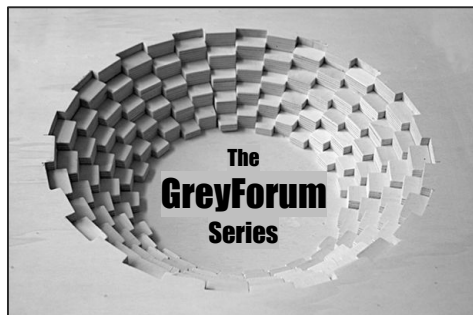
Currently, more than 20 federal agencies – both military and civilian – including the Department of Defense, the Department of Justice, the Department of Health and Human Services, the Department of Education, the Department of the Treasury, the Office of Management and Budget, and the General Services Administration participate in the Federal Strategic Sourcing Initiative (FSSI). The FSSI was created in 2005 by the Department of the Treasury, the Office of Management and Budget, and the General Services Administration to identify products and services that can be purchased more efficiently through a centralized acquisition process. FSSI agencies also provide centralized acquisition functions for a variety of products to streamline efficiency and reduce costs to the federal government.



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

"The GreyForum is a series of onsite and online courses, seminars, and workshops where grey literature provides common ground for information professionals in the process of knowledge transfer"

Wikipedia for Grey Resources: Instruction and Training Skills



This onsite forum and training is organized by GreyNet International together with the African Studies Centre Leiden (Netherlands) and is sponsored by the Federal Library and Information Network of the Library of Congress (United States).

Wikipedia for Grey Resources: Instruction and Training Skills



GreyForum 5.1

Workshop

Wikipedia for Grey Resources:
Instruction and Training Skills

Date

Monday, 2 May 2016

Time

9:00am – 4:30pm

Location

Adams Building, Room LA 151
Library of Congress, 110 2ND St., SE
Washington, DC 20540 USA

Metro/Directions

Capitol South (Orange & Blue Lines)

Sponsor

FEDLINK
Grey Literature Working Group

Information

Please call +1 (202) 707-4813
TTY +1 (202) 707-4995 Request ADA
Accommodations five business days
in advance at ADA@loc.gov

Registration

This program is FREE, however
[ONLINE REGISTRATION](#) is required.

Cancellations

Please notify FEDLINK at +1 (202)
707-4848 if you cannot attend to
allow those on the waiting list an
opportunity to register.

About the Program, Instructors, and Trainees-in-spec

Wikipedia : Successful Against All Odds

Jos Damen, African Studies Centre, ASC Leiden (Netherlands)

GreyGuide : A Repository and Portal for Grey Resources

Dominic Farace, GreyNet International (Netherlands)

Wiki's in Progress : Case Studies from Start to Never Finished

Jos Damen, Wikipedian in Residence
Dominic Farace, Wikipedian in the Sandbox

The program and training is geared to librarians, archivists, researchers, and other information professionals in government, academics, and business involved in raising public awareness to their information resources. Presentations in the morning session are held in a plenary setting, while the afternoon training session is conducted with the aid of computers.

Program Outline

09:00 Arrival, Welcome	13:00 Lunch
09:30 Wikipedia 101	14:00 Training Session
11:00 Grey Resources	15:30 Take Away
12:00 Examples and Cases	16:30 Close

About the GreyForum Series www.greynet.org/greyforumseries.html

GreyNet

Author Information

Biagioni, Stefania

26

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 (ETRD) 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

Carlesi, Carlo

26

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 Antarctic - (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 - (ETRD 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

Ćirković, Snježana

41

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

Farace, Dominic

26

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@grey-net.org

Frantzen, Jerry

26

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.

Joseph, Lura E.

35

Lura E. Joseph is an associate professor and librarian at the University of Illinois at Urbana-Champaign. She has Master's degrees in Geology and in Library and Information Studies. She was a petroleum geologist for 15 years, and has been a librarian since 1994. She was Physical Sciences Librarian at the North Dakota State University, and was the Geology Librarian at University of Illinois from 2001 until the library closed in 2012. She is currently a Content Access and Research Services Librarian. Email: luraj@illinois.edu

Marín Arraiza, Paloma
5, 10

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

Motta, Gabriele
15

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.

Plank, Margret
5

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

Puccinelli, Roberto
15

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
15

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
15

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

Stock, Christiane
26

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



An International Journal on Grey Literature

Subscription Order Form

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

<i>TGJ Volume 12, 2016</i>	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

Notes for Contributors

Non-Exclusive Rights Agreement

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

Submission Requirements

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

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

REFERENCE GUIDELINES

General

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

Specific

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

Examples

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

Crowe, J., G. Hodge, and D. Redmond (2010), Grey Literature Repositories: Tools for NGOs involved in public health activities in developing countries. – In: Grey Literature in Library and Information Studies, Chapter 13, pp. 199-214. – ISBN 978-3-598-11793-0

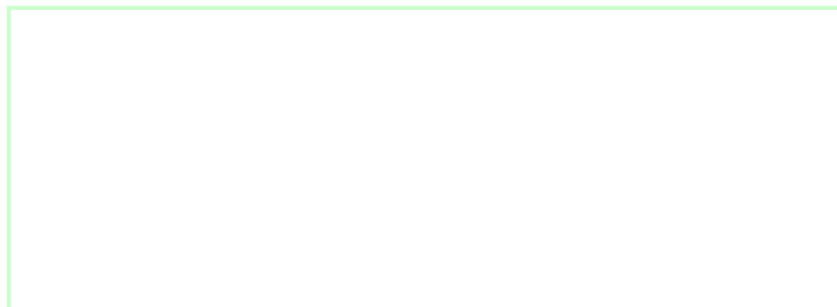
DCMI, Dublin Core Metadata Initiative Home Page http://purl.oclc.org/metadata/dublin_core/

Review Process

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

Journal Publication and Article Deposit

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



An International Journal on Grey Literature
'MINING TEXTUAL AND NON-TEXTUAL DATA SOURCES'
TGJ Volume 12, Number 1 2016

Move beyond text – How TIB manages the digital assets researchers generate	5
<i>Margret Plank and Paloma Marín Arraiza (Germany)</i>	
Scientific Audiovisual Materials And Linked Open Data: The TIB Perspective	10
<i>Paloma Marín Arraiza (Germany)</i>	
Extracting value from grey literature: processes and technologies for aggregating and analyzing the hidden “big data” treasure of organizations	15
<i>Gabriele Motta, Roberto Puccinelli, Lisa Reggiani, and Massimiliano Saccone (Italy)</i>	
Leveraging Grey Literature – Capitalizing on Value and the Return on Investment: A Cumulative Case Study	26
<i>Dominic Farace and Jerry Frantzen (Netherlands); Stefania Biagioni and Carlo Carlesi (Italy); Christiane Stock (France)</i>	
Geologic Field Trip Guidebooks: Shades of Grey	35
<i>Lura E. Joseph (United States)</i>	
Grey Literature Sources in Historical Perspective : Content Analysis of Handwritten Notes	41
<i>Snježana Ćirković (Serbia)</i>	

Colophon.....	2
Editor's Note.....	4
On The Newsfront	
TIB joined GreyNet as an Institutional Member in 2016.....	47
Info Adverts	
GL18 Program Committee – Eighteenth International Conference on Grey Literature.....	48
GL18 Call for Papers – ‘Leveraging Diversity in Grey Literature’.....	49
FEDLINK, The Federal Library and Information Network - Library of Congress.....	50
GreyForum 5.1 Wikipedia for Grey Resources: Instruction and Training Skills.....	51
TGJ Subscription Form 2016.....	54
Author Information.....	52
Notes for Contributors.....	55