https://doi.org/10.26069/greynet-2026-000.555-tgj

ISSN 1574-1796

An International Journal on Grey Literature

Summer 2011 – TGJ Volume 7, Number 2

'System Approaches to Grey Literature'

www.textrelease.com

GreyNet Grey Literature Network Service

www.greynet.org

The Grey Journal
An International Journal on Grey Literature

COLOPHON

CIP
The Grey Journal (TGJ): An international journal on grey literature / D.J. Farace (Journal Editor); J. Frantzen (Technical Editor) ; GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 7, Number 2, Summer 2011. - *EBSCO Publishing, FLICC-FEDLINK, INIST-CNRS, JST, NTK, and NYAM are corporate authors and associate members of GreyNet International. This serial publication appears three times a year - in spring, summer, and autumn. Each issue is thematic and deals with one or more related topics in the field of grey literature. The Grey Journal appears both in print and electronic formats.*
ISSN 1574-1796 (Print)
ISSN 1574-180X (E-Print/PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:
€125 individual, including postage & handling
€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/glpublications.html

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.

About *GreyNet*

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

Full-Text License Agreement

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

Contents
"System Approaches to Grey Literature"

Integration of an Automatic Indexing System within the Document Flow of a Grey Literature Repository 65
Jindřich Mynarz and Ctibor Škuta (Czech Republic)

An Analysis of Current Grey Literature Document Typology 72
Petra Pejšová (Czech Republic) and Marcus Vaska (Canada)

A Profile of Italian Working Papers in RePEc 81
Rosa Di Cesare, Daniela Luzi, Marta Ricci and Roberta Ruggieri (Italy)

From OpenSIGLE to OpenGrey : Changes and Continuity 93
Christiane Stock and Nathalie Henrot (France)

GL Transparency: Through a Glass, Clearly 99
Keith G. Jeffery (United Kingdom) and Anne Asserson, (Norway)

Invenio: A Modern Digital Library System for Grey Literature 105
Jérôme Caffaro and Samuele Kaplun (Switzerland)

Colophon 62
Editor's Note 64

Advertisement

Refdoc.fr – Institute for Scientific and Technical Information 92
De Gruyter Saur, *Grey Literature in Library and Information Studies* 98
J-STAGE – Japan Science and Technology Agency 113

On the News Front

New Password Protected Access to LIS Serials 109
Thirteenth International Conference on Grey Literature – Conference Program 110
The Grey Circuit, From Social Networking to Wealth Creation – GL13 Registration 112

About the Authors 114
Notes for Contributors 115

EDITOR'S NOTE

During the Twelfth International Conference on Grey Literature, I had the opportunity to present for the first time in public forum a back office report on the workings of TextRelease and its Grey Literature Network Service. That comprehensive report included The Grey Journal, which has become an integral component of GreyNet International.

GreyNet was first founded in 1992 and re-launched in 2003 under the leadership and direction of TextRelease. Since its relaunch, it has developed into an international network capable of serving various sectors of government, academics, business and industry as well as subject based communities producing, processing, and distributing grey literature. The mission of GreyNet, which is dedicated to research, publication, open access, and education in the field of grey literature, requires now more than ever an infrastructure commensurate to its real potential.

In order to further expand and develop future capabilities for the Grey Literature Network Service, an infrastructure other than that of a sole proprietorship is essential. While TextRelease provided a basis for GreyNet's relaunch in 2003, it alone cannot render the needed capital and investment for GreyNet's potential to develop and expand on the global information landscape. The ideal organization would be focused internationally and have genuine interest in the field of grey literature both in digital and print formats.

Following the example of other international organizations, the vision and direction for GreyNet would do well to model upon an infrastructure of that of an association. To this end, TextRelease seeks to cooperate with leading institutions in the transfer and registration of GreyNet as a new legal entity i.e. an international association with a designated seat of governance.

Since that first presentation in 2010, new developments have given cause for a revised business report that will be presented in part during GreyNet's 2011 Awards Banquet. On this occasion recognized stakeholders in the future of GreyNet will be present and accounted for.

Dominic Farace, Journal Editor
journal@greynet.org

Integration of an Automatic Indexing System within the Document Flow of a Grey Literature Repository *

Jindřich Mynarz and Ctibor Škuta *(Czech Republic)*

Abstract
The Web empowered the authors of grey literature to publish their work on their own. In case of self-published works their author is also their indexer. And because not many of the grey literature authors are professional indexers, this may result in poor or no indexing.
Even though the Web made publishing easier, indexing is still hard. Nevertheless, we believe that the web technologies and machine learning algorithms may help to reduce the cognitive overhead involved in indexing, and make it eventually as easy as publishing on the Web is.
To help overcome the issue of quality and consistency of subject indexing automatic indexing systems can be used. Given enough full-texts already equipped with the terms from the controlled vocabulary that is to be used, machine learning algorithms can be employed.
Our aim is to provide human-competitive automatic indexing to authors and producers of grey literature. We demonstrate how an automatic indexing system based on machine learning can be integrated into the document flow in an open source digital repository of grey literature. We build upon open source tools and a controlled subject headings vocabulary available in an open standard format.
We will be using Maui Indexer as an automatic indexing system, CDS Invenio as a digital repository software, and Polythematic Structured Subject Heading System (PSH) as a knowledge organisation system. Both Maui Indexer and CDS Invenio are open source, and CDS Invenio's modular architecture makes it possible to extend it with new functionality. Maui Indexer works with controlled vocabularies expressed in Simple Knowledge Organisation System format in which the PSH is available.
From these components combined we will try to put together a solution for automatic indexing aimed at grey literature in the Czech language environment. Maui Indexer is domain and language independent so it is possible to adapt it for the field of Czech grey literature. The document samples we will test on will come from the National Repository of Grey Literature which is maintained by the National Technical Library of Czech Republic.
In the end, we will discuss integration of the automatic indexing component from the user perspective and sketch out how the user can interact with it through the user interface. We will also provide details around the actual implementation of the proposed system. The conclusion will deal with the evaluation of benefits of the implemented system for grey literature authors.

Introduction

The Web offers a publishing model that empowers masses of users to publish their works on their own. As it was stated in the previous literature, grey literature *"does not imply any qualification, is merely a characterisation of the distribution mode"*[3], and the Web can be considered as the single most significant distribution mechanism for grey literature. The sheer volume of documents available on-line constitutes a significant part of the grey literature publishing landscape. The Web has made *self-publishing* easier by lowering both the financial hurdles and the amount of know-how necessary for the publication process and thus it enabled a kind of *"do it yourself"* publishing.
While this is a tremendous benefit without which the *open access* movement would not have been established as firmly as it was, there are also drawbacks to it. By contrast to this mode of making documents accessible, the traditional publication models have procedures in place that go along with publishing that are not replicated well in grey literature publishing on the Web. The part of traditional publishing process that can be neglected in publishing on the Web is *subject indexing*. While it is now for the most part clear how to publish documents on the Web, the approaches to subject indexing are less established and available for use to non-professional users. This topic will be discussed in our paper and our focus will be to show how *self-indexing* of documents published on the Web via a digital repository can be accomplished; much in the same way users were endowed with the ability to *self-publish* their works.

* First published in the GL12 Conference Proceedings, February 2011

Indexing of Grey Literature

Grey literature is characterized by a way of publishing that outputs documents with limited visibility. It may be hard to find such documents because they are distributed in a way that does not use established document access mechanisms, such as commercial databases and the like. This aspect of grey literature makes it difficult to be searched for either through libraries or web search engines. Also, the field of grey literature is closed tied to the *open access* way of publishing. However, if a document *cannot be found* there is no use of it being released in the *open access* way.

As we will argue, additional *subject indexing* terms can make grey literature documents searchable in a meaningful way so that they eventually become more prominent in the search results of library or web search engines. Subject indexing can be seen as an essential requirement to make documents findable [8], even though there are powerful search engines that enable documents to be found even without carefully crafted indexing. It can help in making grey literature documents more visible. Moreover, there is a consensus that it is necessary for useful navigation interfaces that can be built on top of digital document repositories (e.g., a faceted navigation) [7].

Subject indexing metadata enriches documents with *affordances* that allow to do more with them. It can support navigation interfaces that make it possible to browse the document collection in a useful manner based on *navigation paths* taken from the structure of the knowledge organisation system used for indexing. In this way the connections within the subject indexing system, such as hierarchical or associative relationships, may be harnessed as a *"map"* to the document collection. The subject indexing places the documents in a logical space constructed by the knowledge organisation system's structure and allows the user to browse the documents organized in such way by following the relationships between indexing terms. Every subject indexing term that is assigned to a document constitutes an *entry point* through which the document can be found and accessed.

Thus, we see subject indexing as an important enrichment to the document that enables to build interfaces for document repositories that can be navigated in a meaningful way. In this paper, we are interested in subject indexing done by the grey literature authors, and we have to admit that there are barriers that can make subject indexing a difficult task for them.

Grey literature is often published directly by the documents' authors, which may imply that, if it contains any subject indexing terms at all, it is the indexing done by the *authors themselves*. Subject indexing may be difficult for non-professionals and therefore this situation may result in no or poor indexing.

The established best practise for subject indexing is to use a *controlled vocabulary*. Authors know best about the contents of their works but they might not be familiar with the controlled vocabulary that they are supposed to use to express their works' content. They may not know how best to use the subject indexing system to describe their documents and therefore it constitutes a barrier for them, because first they have to learn how to use it.

This is why professional indexers are able to produce better indexing - they know how to *use* the subject indexing terms, especially with respect to the whole document collection. This background knowledge of the indexing system and the document corpus is a fundamental requirement for high-quality indexing.

In automatic indexing this background knowledge is in a sense captured in the *indexing model* on which the automatic system operates. And thus it turns out that, if configured properly, automatic indexing might come up with subject terms of reasonable quality and consistency. In the following sections we will describe how to reach that goal, while we will deal in detail with an application of automatic indexing to grey literature.

Automatic Indexing

In this paper we will investigate the option of non-professional *self-indexing* of grey literature on the example of an automatic indexing system for a digital document repository. We propose a semi-automatic indexing system that incorporates human feedback for the final selection of indexing terms. The system suggests a set of pre-selected indexing terms from a controlled vocabulary that may be used to describe document's content. These terms can be refined in the next step by the user interacting with a selection via interface that enables to remove or add new indexing terms.

The main help of our proposed approach is to lessen the cognitive overhead involved in intellectual indexing. Rather than completely automating the process of subject indexing we decided to use the automation for suggestions of indexing terms that can be amended and validated by the human user.

In this way we strive to provide grey literature authors with a tool that makes it possible for them to come up with non-professional indexing that is of high quality and consistency. The

reason for such a goal is that the inconsistency of *user-generated indexing*, often done with freely created keywords, is one of its main drawbacks. This is what we try to alleviate by preprocessing the document and suggesting indexing terms in an automated manner. Also, we argue that this approach might help to increase the *scalability* of subject indexing.

The intellectual indexing carried on by professional indexers does not *scale*. Given that the size of the grey literature published on the Web is ever-increasing, there is a need for indexing system that is able to scale. The traditional solutions for subject indexing based on manual examination of each processed document do not provide a way to scale them up and therefore they are not the best answer for the requirements of the current grey literature publishing.

Scaling is a difficult problem in any situation and we do not strive to provide a definitive answer on how to scale up subject indexing in general or even in the case of grey literature. We propose that this issue can be alleviated by the use of automatic indexing. This solution scales because the time that is being used for subject indexing is *machine time* instead of human time. The processes that are carried out automatically with computers can be scaled up by assigning more computational power to them. Although in fact, our focus is not to achieve full automation of the indexing process so that it may be scaled on demand, but rather *to scale up the number of people* that are able to do the indexing while retaining a reasonably high quality.

Now that we have described the aim we try to achieve with automatic indexing, we will continue with a basic description of what automatic indexing is. It is a process of assigning indexing terms to a document in an automated fashion. It can use techniques based on analysis of language corpora. A common form of automatic indexing relies on "simple" statistical analysis of the full-text. In the case of our indexing system, we have used automatic *term assignment* that selects terms from a controlled vocabulary for a particular document based on the analysis of the document's content. We have employed a *machine learning* approach that is based on computing conditional probabilities.

The approach we have chosen employs *supervised machine learning* that gathers feedback from users of the indexing system. Machine learning modifies the automatic indexing functionality with regard to the set of training data on which it "learns" how to do good indexing. The approach of supervised learning adjusts the indexing algorithm based on the newly acquired information about the way the indexing system is being used. Each time user approves of a set of automatically generated subject terms this decision is fed back into the indexing model and makes it more aligned with the specifics of the document collection.

One of the fundamental requirements of automatic indexing is the access to the *full-text* of the examined document. However, this is not a problem in the use case we have described. If the indexing of a document is done by its author, the access to the document's full-text is granted.

Automatic indexing consists of a sequence of processes. During the course of automatic indexing the full-text is processed by a number of procedures that are collectively referred to as the machine processing *pipeline*. The full-text is sent through a sequence of processes that take the text as their input and pass their output to the process that is directly after them in the pipeline's sequence.

In our case we start with a procedure that yields a *plain-text* of the document in question which might have been in another format, such as PDF or MS Word. Once we have acquired a plain-text embodying the content of the document a series of normalizations and helper procedures are run on it.

One of them removes the *stop words* - the words that do not affect the meaning of the document, such as prepositions or conjunctions. The system contains a list of stop words that are automatically excluded from further processing.

Another common technique that we take advantage of is *stemming* which reduces words to their root forms. In this way, we get rid of inflections, plural suffixes, and other characters that differentiate among the derivatives of the same root form. This method is supposed to collapse the different forms that refer to the same meaning to one word form so that a more effective computation can be done with it.

After these pre-processing steps automatic indexer carries on with its main suite of functions; it analyses the full-text and outputs a set of suggested indexing terms that can be assigned to the processed document. Since this paper is not about automatic indexing itself, but rather its application for grey literature, we will not discuss this part in detail and instead, we will move to the actual implementation of the indexing system.

Implementation

After having described the field of automatic indexing in general we will now proceed to provide an overview of the way we have implemented the automatic indexing system and put the methods of automatic indexing for grey literature into practice.

The guiding principle of our implementation was *re-use* of existing components, which we combined together in a document processing pipeline, or extended them in the cases where there was a need for it. This way of development would not have been possible if the parts we were building with had not provided access to their source code. Hence, their *open source* nature enabled us to modify them and extend their functionality. The combination of the constituent parts was possible due to their modular architecture that enabled them to be joined in a chain of processing procedures, which are applied on the examined document.

Not only the software that we have used in the automatic indexing system was open source, the data is communicated in this system in *open formats*. To illustrate this point, the subject headings system we have used was already available in RDF[1] data format expressed with SKOS,[2] an established standard for representing knowledge organisation system, such as thesauri, subject headings systems, or systematic classifications. It has gone through the standardization process and has reached the status of a recommendation of the World Wide Web Consortium.[3] This open standard is well supported by the indexer we have used, which eliminated the necessity of data conversion to a suitable format.

In order to manage the flow of control in the system a unifying data communication format is used. For this purpose we have adopted *JSON*, a light-weight data communication format.[4] The parts of the system exchange short JSON messages to pass the data needed for the indexing process to another part of the system. In this way we have harnessed another standard format to glue the components of the system together.

We wanted to preserve high modularity of the individual components in the system as a whole as well. Therefore we have exposed most of the functionality of the resulting system as a *web service*, which encourages loose coupling and re-use of the system's parts.

Now that we have described the overall architecture and design of the system we will move to the discussion of the individual parts. In the section that follows we will present an overview of the components that are involved in the automatic indexing system we have built.

Components

We will briefly describe each of the components that together make up for the whole automatic subject indexing system. These are not strictly limited to *software* but they also include *data* that is used in the process of subject indexing – the subject headings system from which the indexing terms are drawn, and the full-text corpus which serves for machine learning algorithms.

According to the design goal we have stated previously, the automatic indexer pipeline is composed mainly of already existing applications that we have re-used for this purpose. The parts that are new serve to connect the re-purposed components. We have written the *"glue code"* that ties the parts that are used in the process of automatic indexing. In order to connect all the components together and set up the indexer for processing Czech language, only a few additional functions had to be implemented.

Subject Headings System

One of the core components of the system is the controlled vocabulary of subject headings we have used. It provides indexing terms that are assigned to documents, which enables to maintain a degree of indexing's consistency by referring only to indexing terms that are *authorized* by the subject headings system. The use of a controlled vocabulary implies that the suggested indexing terms are more consistent compared to the keyword extraction techniques.

The subject headings system we have used is the *Polythematic Structured Subject Heading System* (further abbreviated as PSH).[5] PSH is a bilingual Czech-English controlled vocabulary maintained and used at the *National Technical Library*.[6] It is a universal system and it consists of headings describing all major aspects of human knowledge. Its structure is similar to thesauri with hierarchical, associative, and equivalence relationships. PSH is primarily expressed in the *MARC 21 Format for Authority Data*,[7] but it was also converted to RDF data format, expressed with SKOS, which is more suitable for the automatic indexer we have employed.

Digital Repository

The *"host environment"* in which the system of automatic indexing is built in is the digital repository software. The software we have used for this purpose is *CDS Invenio*.[8] This

software's modular architecture enabled us to extend its functionality with a new plug-in that does the automatic indexing.
Invenio processes newly submitted documents in a series of step that are referred to as the document workflow. We have inserted the automatic indexing pipeline into the workflow so that a document can go through this additional suite of procedures to be enriched with subject indexing terms.
The user interface of the automatic indexing system is included in the repository. To achieve this level of integration not only we had to extend the functionality of the modified software but also alter the presentation interface to enable the user to access the added new functionality. This was possible due to the clear separation of the code responsible for the repository's core functions and the templates that build up the interface the user is interacting with.

Automatic Indexer

The component that is responsible for the main function of the system is the automatic indexer. We have chosen to use *Maui Indexer*.[9] The author of this software claims that it produces *human-competitive indexing*[4] and it seems correct from the results of the comparative studies done with this indexer.[10] This assertion is based on the implicit presumption that the subject terms assigned by humans are the standard with which the quality of automatic indexing is compared. The indexing produced by Maui Indexer is thus comparable to human indexing both in terms of quality and consistency, which is precisely the result we are looking for in our automatic indexing system.
The software is described as being independent of the domain and language for which it is used. However, to achieve the best precision some adjustments need to be done. Because of the language of the document collection, for which the automatic indexer was intended, we wanted to adapt Maui Indexer for the Czech language. The modifications involved changing of the indexer's parts: the list of stop words and the stemmer code.
In the ideal case, stop words would be based on the corpus of documents that we want to index. We chose to use the *Czech National Corpus*[11] instead to create a list of the most frequent words that may be used as stop words. Czech National Corpus is a vast document collection reflecting the contemporary written Czech [2]. Thus, it served well to establish a good baseline with respect to our document collection.
To reduce words in an analysed full-text to their root forms we have taken over the *aggressive Czech stemmer* [1], which we have adapted in a way so that it can be plugged in the Maui Indexer's source code.[12] The aggressive nature of the stemmer is based on the approach it takes to stemming non-root word forms. It addresses the morphological characteristics of the Czech language to normalize the irregularities of inflection, consonant alterations and the like. However, in some situations, it may remove characters that are necessary for the distinction of the word sense and thus create the same root form from multiple words that do not share such a root. This feature may compromise the quality of resulting indexing and it is the reason why we consider it as an immediate target for further refinement of our indexing solution.

Text Corpus

The procedure built by combining the previously mentioned components was applied to a text corpus of the document collection of grey literature documents stored and maintained in our digital repository. In our case, we have applied the automatic indexing system we have built to the *National Repository of Grey Literature*.[13] This repository, maintained at the *National Technical Library*, collects grey literature from the network of cooperating partner institutions, ranging from the institutes of the Academy of Sciences of Czech Republic to public universities [6].
The contents of the documents included in this repository are mostly in Czech. This was the primary drive behind the decision to enhance the functionality of the automatic indexer towards the Czech language. We have also taken into account that the contents of the repository are produced in collaboration with the partner institutions. Its long-time goal is to have the co-operating institutions produce the document's descriptive metadata, including subject indexing, on their own without a need for central co-ordination. This intention made for an adequate use case of the author-generated subject indexing.

User interface design considerations

While the quality of the underlying procedures is certainly crucial to produce results of a reasonable quality, the part of the indexing system that has a comparable significance is the *user interface*. The importance of the user interface design stems from the necessity of user feedback. The way how the users provide feedback on the automatically generated set of indexing terms needs to be designed carefully to take advantage of the author's knowledge

about the indexed document. The resulting design has some notable features that may have a significant influence on the user experience with the indexing system.

We have decided to provide the automatic indexing as an *opt-in* procedure, which means that the user has to actively declare that the document entered into the repository should be processed with the automatic indexer. If the user checks in a box for automatic indexing, during the next step in the document workflow there will be an additional screen containing the suggested indexing terms and the functionality that allows to modify them.

The primary functionality of the automatic indexing system we have developed is to suggest a list of subject headings that in some way describe the processed document. Its objective is to facilitate non-professional indexing while maintaining a high level of consistency. It is not meant to serve as a *replacement* for the person doing the indexing, but in fact, it is a start for the indexing process controlled by the human user.

We still see the main value being added to the processed documents by the user of the system rather than generated by the system itself. For this purpose we have enhanced the user interface with helper functions that are meant to facilitate more effective indexing.

One of these functions is the *auto-complete* feature. To bridge the gap between the language of the user and the language of the knowledge organisation system used for indexing the user interface displays a list of suggested headings based on a heading's fragment supplied by the user.

To assist the user in deciding on the subject heading's adequacy we have added a utility that shows citations of the example documents indexed with the subject heading user considers to use. When user highlights certain subject heading a short list of links to the documents which have the same heading attached is presented in the interface. In this way, the user can *learn by example* to consider the applicability of a given subject heading for a particular document.

We had to make alterations to the *search interface* as well. There would be no value in added indexing terms if one could not use them to search and navigate the document collection in which they are used. To reflect this added structure a change to the user interface needs to be made so that it is possible to harness the indexing terms to access documents. In our case, responding to this requirement consisted in adding a new search index built for the subject headings and appending a new field to the search form to access this index.

Future Possibilities and Challenges

The work we have done with the automatic indexing system for grey literature is by no means finished and we are aware that there are further possibilities for improvement and challenges that have to be solved to deliver a better system. In the specific use case we have described in this paper there are certain aspects of the automatic indexing that are worth underscoring.

It is important to note that there is no use in adding subject indexing terms if such metadata cannot be harnessed via the search or navigation interface for the document collection. If the indexing is not reflected in end-user interfaces it does not provide any additional value. Thus, the indexing must be represented in user interfaces to have an impact on the overall functionality of a digital repository. As we have mentioned in the previous section, we have extended the digital repository with a new search field to access the documents through subject headings to address this issue.

We have to take into account that the indexing we are dealing with here is still a *user-generated indexing*, even though it is somehow refined by the system we have implemented. This implies that the resulting indexing terms might need further verification by a professional if we want to have a quality control in place. As we have sketched in the paper, this is the way our system works and will work for the near term future. Nonetheless, our goal is eventually to get rid of this necessity once we will have a higher level of confidence in the non-professional subject indexing that is fed into the repository.

Due to the modular nature of the whole system, there is a plethora of ways how it can be enhanced and developed further. Every part of the document processing pipeline can be considered for an improvement. Our aim is first to focus on the parts that affect the quality of the results of indexing the most. After every change we want to check if it leads to an increase in the system's precision by comparing it with the results the system had on the same document with the previous configuration.

We argue that the system not only benefits grey literature authors and maintainers of digital repositories, but, moreover, it can also benefit the individual components it is made of because it reflects on the way the component is being *used*. This can be applied on the knowledge organisation system from which the indexing terms are drawn. The data about its usage in practise can be crucial for its development and further evolution reflecting the changing needs of the user community and a shift in the way its concepts are perceived.

Conclusions

This project would not have been possible if there were not *open standards* that govern the field of subject indexing. They enabled the re-use of existing components adhering to certain standards and their combination in a novel way. The layer provided by open standards constituted an environment for interoperability and systems built with an open architecture in mind.

All the parts we have put together in this open framework are *open source*. This means they are open to modifications and extensions, and those were necessary for the system to work as a whole. Moreover, due to the their modular character it was possible to switch one part for another or plug in a new component.

The indexing system that came out of this way of development is applied to the grey literature documents. It was designed to reflect the nature of grey literature. We have argued that the situation of subject indexing of grey literature is unsatisfactory and we have expressed our view of the causes for such state. Our motivation was to react to the current conditions and propose an approach that may lead to an improvement of the way subject indexing is done for grey literature.

References

[1] DOLAMIC, Ljiljana; SAVOY, Jacques. Indexing and stemming approaches for the Czech language. *Information Processing & Management*. November 2009, vol. 45, iss. 6, p. 714 – 720. ISSN 0306-4573. DOI 10.1016/j.ipm.2009.06.001.

[2] KUŘERA, Karel. The Czech National Corpus : principles, design, and results. *Literary and Linguistic Computing.* 2002, vol. 17, no. 2, p. 245 – 257. ISSN 0268-1145.

[3] MACKENZIE OWEN, John S. The expanding horizon of Grey Literature. In *Perspectives on the design and transfer of scientific and technical information : proceedings of the 3rd international conference on grey literature*. Amsterdam : TransAtlantic, 1998, pp. 9–13. Also available from WWW: <http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/glpaper.htm>.

[4] MEDELYAN, Olena. *Human-competitive automatic topic indexing*. Waikato, 2009. 214 p. Dissertation thesis (PhD.). University of Waikato, Department of computer science, 2009. Also available from WWW: <http://www.cs.waikato.ac.nz/~olena/publications/olena_medelyan_phd_thesis_July2009.pdf>.

[5] PEPE, A. [et al.]. CERN Document Server Software : the integrated digital library. In DOBREVA, Milena; ENGELEN, Jan (eds.). 9th *ICCC International Conference on Electronic Publishing : from author to reader : challenges for the digital content chain.* Leuven : Peeters, 2005. ISBN 90-429-1645-1.

[6] PEJŠOVÁ, Petra (ed.). *Grey literature repositories*. Zlín : VeRBuM, 2010, 156 p. ISBN 978-80-904273-6-5.

[7] RIBEIRO, Fernanda. Subject indexing and authority control in archives : the need for subject indexing in archives and for an indexing policy using controlled language. *Journal of the Society of Archivists.* April 1996, vol. 17, iss. 1, p. 27 – 65. ISSN 0037-9816.

[8] SYKES, Jan. *The value of indexing : a white paper prepared for Factiva, a Dow Jones and Reuters Company* [online]. February 2001 [cit. 2010-12-02]. Available from WWW: <http://4info-management.com/pdf/indexingwhitepaper.pdf>.

[9] VLACHIDIS, Andreas [et al.]. *Excavating grey literature : a case study on rich indexing of archaeological documents by the use of natural language processing techniques and knowledge based resources* [online]. 2009 [cit. 2009-09-30]. Preprint for ISKO UK 2009 : Content Architecture: Exploiting and Managing Diverse Resources. Available from WWW: <http://www.iskouk.org/conf2009/papers/vlachidis_ISKOUK2009.pdf>.

Notes

[1]Resource Description Framework. <http://www.w3.org/TR/rdf-concepts/>.
[2]Simple Knowledge Organisation System. <http://www.w3.org/TR/skos-reference/>.
[3]http://www.w3.org/
[4]http://www.json.org/
[5]The on-line version is available at http://psh.ntkcz.cz/skos/
[6]http://www.techlib.cz/en/
[7]http://www.loc.gov/marc/bibliographic/
[8]The project's website is at http://invenio-software.org/ and our installation of Invenio is available at http://invenio.ntkcz.cz/.
[9]http://code.google.com/p/maui-indexer/
[10]Examples can be found at http://code.google.com/p/maui-indexer/wiki/Examples
[11]http://ucnk.ff.cuni.cz/english/index.php
[12]The stemmer we have used is available at http://members.unine.ch/jacques.savoy/clef/CzechStemmerAgressive.txt.
[13]http://nrgl.techlib.cz

An Analysis of Current Grey Literature Document Typology*

Petra Pejšová *(Czech Republic)* and Marcus Vaska *(Canada)*

Abstract
This analysis is based on the classification of the international systems GreyNet, (the Grey Literature Network Service), OpenSIGLE, (the System for Information on Grey Literature in Europe), and the Registry of Open Access Repositories (ROAR), as well as focusing on national schemata in the Czech Republic, namely ASEP (Register of Publication Activity of the AS CR), NRGL (National Repository of Grey Literature), and RIV (Information Register of R & D Results). During the analysis of the lists of document types, we have discovered that these typologies contain, besides "real" document types (reports, theses, etc.) other aspects, such as events (arrangement, organization), types of events (conferences, speeches), producers (universities, institutes), processes (translations, output), content (political documents, legal texts), location (domestic, foreign), and format (e-texts, numeric data). However, this approach is not systematic. Therefore, we have decided to create a classification scheme for document types only, and classify other aspects into various groups in order to define them more precisely. The scheme will be processed in a text version as well as schematically in mind maps.
We believe that identifying a specific typology for credible grey literature document types, particularly reports, conference proceedings, and government documents, will assist in the classification of grey literature in the fields of science, research, and education. On the other hand, grey literature also consists of various means of communication, such as telephone calls, meetings, e-mails, blogs, interviews, social networking tools, or discussions in Wiki. It is important to identify only credible document types and not use unverified information that may be unsuitable for scientific work.
The aim of this analysis is therefore to create, define, and implement a current credible grey literature document typology, in order to open discussions in the grey literature community, leading to a means of collecting GL from reputable events and producers rather than relying on social networking tools or Wiki contributions. While the later types of sources can assist researchers, scientists, and teachers with their information-seeking pursuits, documents of this nature need to be evaluated on a regular basis.

Keywords: analysis, classification, documents, grey literature, systems, types, typology

Introduction: Defining Typology in the Grey Literature

> *"Improved access to and sharing of research information is the key to accelerating progress and breakthroughs in any field" (Brian Hitson and Lorrie Johnson, 2009)*

Indeed, continued and concentrated efforts in the pursuit of grey literature has caused a transparency and ability to share documents that only a few years ago would not have been deemed possible. Despite the wealth and variance of forms of grey literature (including multimedia) in institutional repositories, there appears to be a lack of systematically classifying these documents into a universal, standard typology. In fact, Beissel-Durrant (2004) states that the typologies currently in place in the social science literature "do not necessarily categorize research methods in a systematic way, using mutually exclusive categories and hierarchies that are not necessarily complete" (p.2). An exhaustive search is therefore required to eliminate bias, as the presence of bias could potentially undermine the research retrieved.
Traditional monikers assigned to grey literature have labeled this material as mainly consisting of primary sources, focusing on theses, reports, and government publications. This leads to two key characteristics of grey literature resources, namely their ubiquitous nature, and their difficulty in being properly identified (Schöpfel, 2006). While the history of grey literature may have supported this notion, the same does not necessarily hold true today; Hjorland (2006) claims that "each sphere in society has developed its own kinds of documents." Hence, conducting an analysis of current grey literature document typology and subsequently creating and implementing a quality control system to guarantee the credibility of grey literature becomes increasingly important.
The word *typology*, first introduced in the Merriam-Webster dictionary in 1845, has origins that date back to biblical times. Originally defined as "a doctrine …holding that things in

* First published in the GL12 Conference Proceedings, February 2011

Christian belief are prefigured or symbolized by things in the Old Testament", its current association refers to a "study of or analysis or classification based on types or categories" (Merriam-Webster, 2010). Although various disciplines (i.e. anthropology, archaeology, psychology, and in particular linguistics), have a different notion of how typology fits in with their subject area, the idea of classification and organization is the same. This characteristic is significant when it comes to identifying the role that document typology plays in the realm of the grey literature.

Is typology for grey literature really irrelevant? Some researchers claim it just might be, particularly as the grey literature is moving closer to the white (Di Cesare, 2006). As the numerous document types in the *GreyNet* repository can attest to, grey literature truly does transcend boundaries and plays a role in more than merely science, research, and education. When categorizing the various types of grey literature into a single, universal typology, it seems plausible to see the wealth and versatility of grey literature sources as an information neighborhood, "an environment within which practical information seeking and orienteering information seeking, as well as both directed and undirected browsing, can take place." (Burnett, 2000) The representation of grey literature in numerous types and formats can indeed create the appearance of numerous aspects of grey literature that do not appear to hold a common purpose with each other, hence the need for a standard typology for this material to put everything in its place.

Aspects & Analysis of Grey Literature Typologies

In preparation for the analysis of a grey literature document and subsequently its typology, it is necessary to consider the purpose of the material being studied, the aspect of the grey material in question, and perhaps most importantly, "how are grey documents used?" (Schöpfel, 2009, p.7)

This paper, and subsequently the grey literature typology proposed, will focus on the following aspects: document type, event, producer, content, location, format, and periodicity. While theses, dissertations, reports, conference proceedings, working papers, and to a lesser extent, courseware, are considered fundamental types of grey literature (Schöpfel, 2009), other aspects or 'shades' of grey' must also be considered. The purpose of a document typology for grey literature adheres rather closely to Beissel-Durrant's (2004) notion of prioritizing research methods, exemplifying the focus of the research, identifying needs for additional training or research, as well as appropriate classification methods.

Creating a Grey Literature Typology Classification System

We collected a total of 241 terms used to describe grey literature typologies; these terms have been categorized into one of six system typologies as follows:

- 133 terms from GreyNet
- 35 terms from NRGL
- 30 terms from ASEP
- 17 terms from OpenSIGLE
- 14 terms from RIV
- 12 terms from ROAR

After removing duplicates, we obtained a final list of 193 original terms. While no single term was present in all six typologies, two terms, namely theses and research reports, appeared in five of the analyzed system typologies (theses in GreyNet, OpenSIGLE, ASEP, NGRL, ROAR and research reports in GreyNet, RIV, ASEP, NGRL, ROAR), while 28 terms occurred twice, and six terms materialized three times. It is interesting to note that four instances were observed where no common term was found among any of the system typologies studied.

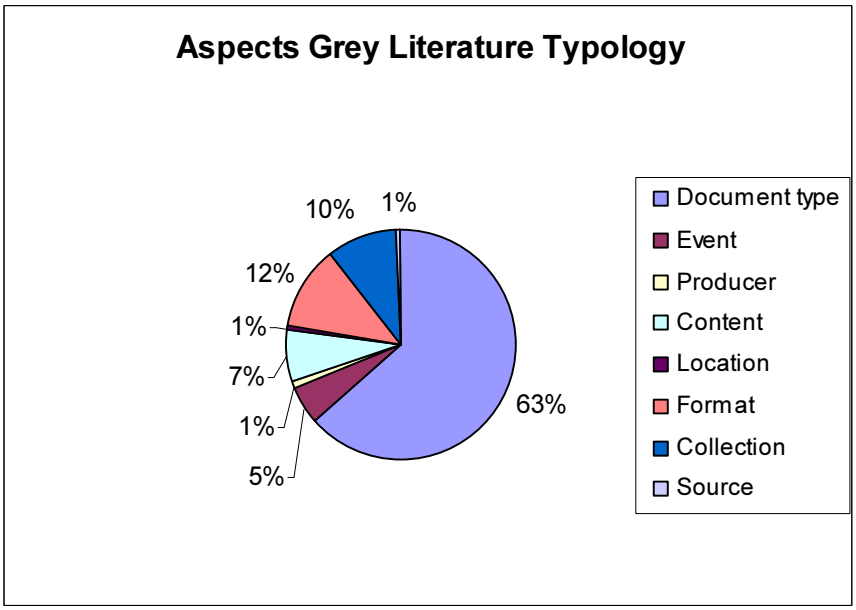
Our examination of the various aspects in grey literature typologies ascertained 121 document types which we subsequently organized into 19 collections. For example, a collection labeled with the broader term ***REPORT*** is comprised of narrower term document types such as annual report, business report, bank report, and so on. The 19 collections identified as narrower term document types are listed in a mind map appearing at the end of this paper (Appendix 1).

In addition to document type, we have proposed six additional aspects of GL typologies in order to more succinctly classify the GL terms we analyzed. These include:

- **Format,** describing the type of presentation (e.g. electronic document, e-text, multimedia, and so on). – 23 terms
- **Content**, referring to the type of information in the document (e.g. computer program description, policy document, product data, and so on) – 14 terms

- **Event,** depicting the occasion on which the document was issued (e.g. conference, workshop, lecture) – 10 terms
- **Producer,** denoting the organization producing the document (e.g. legislation) – 2 terms
- **Location,** indicating the place where the document is situated (e.g. board) – 1 term
- **Source**, representing the source data for the document (e.g. survey) – 1 term

The graph below illustrates the occurrence of the 193 original terms used to describe grey literature amongst the proposed aspects in GL typologies that have been identified.

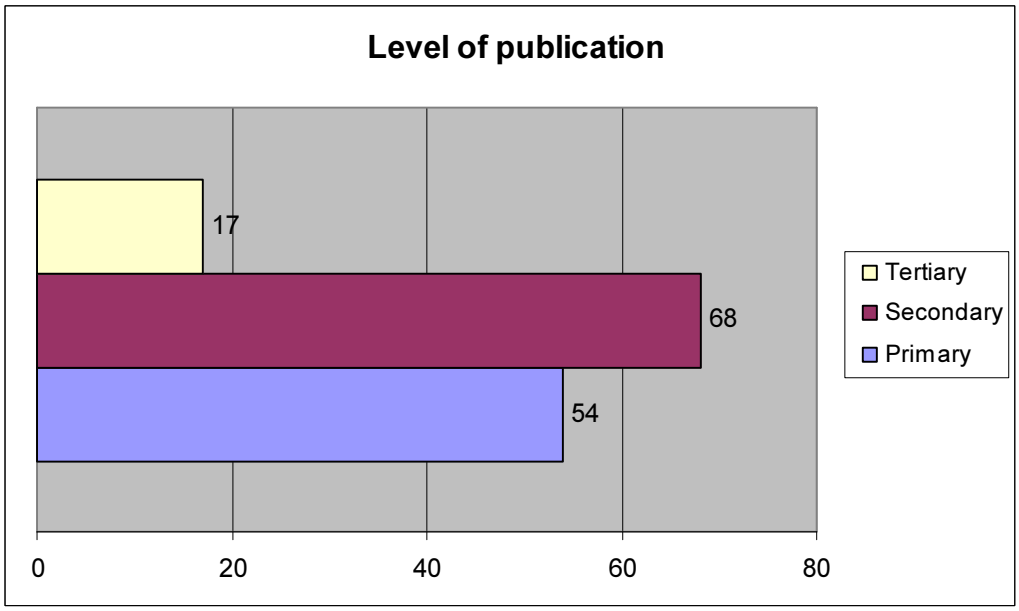


Many of the analyzed terms contain more than one aspect, such as a project information document, (comprised of both and information and a project document), a computer program description, etc. We also identified 44 terms that expressed a similar meaning, requiring us to select preferable terms in order to maintain consistency amongst our classification scheme. For instance, *annual* was deemed the preferred term when describing a yearbook or other such yearly document, while *leaflet* was favored over flyer when classifying items of an advertising nature. However, some terms were difficult to analyze, due to lack of description or unknown reasons as to why they were introduced. This was particularly indicative of website reviews, where questions arose whether the material was a review of a website or rather a review presented on a website. We therefore require and look forward to comments and feedback from the grey literature community on these matters.

Distribution terms to broader (Collections) and narrower terms document types is a basis for creating a new and well-structured grey literature typology. The National Technical Library in Prague will prepare a first draft version that will be available for feedback from the grey literature community. After incorporating the comments received, a subsequent step will implement the aspects identified in the analysis as an optional description of the document types. These optional descriptions are necessary to precisely express the document type without having to constantly introduce new types of documents, thereby unduly expanding the typology of terms that are used only for specialized areas. This approach is necessary for creating a functional general grey literature typology.

In addition, we identified three levels of publication in our analysis, according to the Library of Congress specifications for identifying primary, secondary, and tertiary sources, and Hjorland's (2006) revised version of the UNISIST typology model used for classifying scientific and technical documents :

- 54 primary
- 68 secondary
- 17 tertiary



At this point, this data is being used solely for informative purposes; there are no plans to include levels of publication for creating the first version of a new grey literature typology.

Implementing a Quality Control System to Guarantee Credibility of Grey Literature

While the existence of any electronic media or non-traditional resource should be analyzed for its acceptance as a grey literature document, the adage of effective searching, whereby key websites and the Internet in general are consulted, needs to be considered as well (Giustini, 2010). Social media, blogs, phone, fax, e-mail; all of these forms of grey literature can greatly enhance the searching process. The recent H1N1 epidemic indicated the importance that common social networking sites, primarily Twitter, played in disseminating information in a timely, virtually instantaneous manner. While social networking tools or Wiki contributions should not be relied on exclusively, "the key is not to rule anything out, not even tweets" (Giustini, 2010).

Although the rapid "information-at-your-fingertips" approach is impressive, there is a danger of accepting unverifiable information as fact without further investigation. A case in point is an article that appeared in the United Kingdom's *Daily Mail* newspaper earlier this year (Rose, 2010). Climate change is a global issue, with numerous organizations, such as the World Wildlife Fund (WWF) and the Intergovernmental Panel on Climate Change (IPCC) taking a firm stance and commitment to reducing our carbon footprint and preserving the environment for generations to come. In 2007, a report produced by the IPCC, which was subsequently awarded the Nobel Prize, claimed that the Himalayan glaciers would melt by 2035 (Rose, 2010). Despite repeated objections by some glacial experts as to the accuracy and authenticity of this report, the claim was not refuted until January 23, 2010, when Dr. Murai Lai, a scientist at IPCC, admitted that the statement was "included purely to put political pressure on world leaders." Further investigation revealed that no peer-reviewed scientific research had been carried out to support the Himalayan glacier melting assertion; "the 2035 melting date seems to have been plucked from thin air", and was due to an arithmetical error by the WWF. Unfortunately, this error not only damaged the reputation of the IPCC, it also questioned the quality and qualifications of those producing grey literature, particularly since this material is often not peer-reviewed.

Content contained within open access repositories still prefers theses and dissertations as key primary material. Other grey literature documents, such as conference papers, reports, even works in progress, are slowly buckling the trend; this is comforting, especially since Schöpfel (2009) reports in his study that "100% of the institutional archives give access to grey material." (p.15)

Lack of bibliographic control is the primary reason why grey literature can be difficult to locate; it must be easily retrievable in order to be useful. While most repositories, including GreyNet and OpenSIGLE, are making conscientious efforts to classify the material they store, there are others that still do not do so.

Despite the dissolving of EAGLE in 2005, and SIGLE's dormancy, some researchers believe that each association will classify its documents in-house, without adhering to any bibliographic standards. However, this may not necessarily be as troubling as originally

thought: indeed the rapid expanse of the Internet has enabled increasing numbers of grey literature documents to be made available to the public for the first time, which puts additional pressure on cataloguing this material in a uniform matter. Nevertheless, the Web has also created greater awareness that these types of material exist and can be retrieved. The trade-off between access and awareness is a never-ending challenge.

Evaluating Grey Literature Document Types

Presently, open access repositories of grey literature are maintained and/or funded by either an academic institution (typically a University), or by a pubic research association (Schöpfel, 2009). As the analysis of the following types of documents found in repositories will indicate, a majority of these storehouses of information centre on more than one domain, often playing a multidisciplinary role. Barely seven years old, the notion of Open Access, whereby scholarly output is freely available to the general public, has taken academia by storm. Focusing on institutional repositories and various policies regarding unrestricted access, "the GreyNet community intensified its research activities on the impact of the open access movement on the grey literature." (Schöpfel, 2009, p.4)

Numerous studies have been launched in an effort to determine why research is primarily disseminated via a report, thesis, or conference proceeding, despite the various types and formats that exist today. Schöpfel (2006) ponders this dilemma and provides his own reasons: "research results are often more detailed in reports, doctoral theses, and conference proceedings than in journals…they are distributed in these forms up to 12 or even 18 months before being published elsewhere. " (p. 68)

GreyNet

While the aim of this paper is to analyze, identify, and create a typology for credible types of grey literature, namely reports, conference proceedings, and government documents, the impact of virtual communities cannot be discounted. The notion of creating awareness is often commented upon in papers on grey literature, and social networking tools or Wiki collaborations certainly have a role to play in this pursuit. Announcements, one of the aspects categorized in the *GreyNet* repository is considered to be a common activity in online communication, with some even saying that these information updates "play a significant role in the informational economics of the community" (Burnett, 2000). Information is not just given away; it becomes invaluable to the person seeking it.

Founded in 1992, The Grey Literature Network Service, or GreyNet as it is commonly called, is "dedicated to research, publication, open access, and education in the field of grey literature" (GreyNet, 2010). For nearly two decades, this organization has strived towards seeking, identifying and disseminating grey literature to as wide an audience as possible. Recent technological advances and increasing acceptance and adherence to the Open Access movement have strengthened the awareness of the importance of grey literature among several disciplines. GreyNet has certainly achieved a number of milestones in a relatively short period of time; its goal of facilitating "dialog and communication between persons and organizations in the field of grey literature" will undoubtedly grow exponentially in the coming years.

In 2004, GreyNet developed a GL Survey, whereby visitors to the site were invited to contribute to a list of document types that they felt best constituted grey literature, the purpose of which was to best describe the type of document it embodies. As a result, 133 different document types in grey literature have been identified.

OpenSIGLE

OpenSIGLE, the System for Information on Grey Literature, functions as a repository for the collection of scientific, technical, economic, and humanities documents produced across Europe. The 13 member countries include Belgium, the Czech Republic, France, Germany, Hungary, Italy, Latvia, Luxembourg, Portugal, Russia, Slovakia, Spain, and the United Kingdom (OpenSIGLE, 2010). The partnership between GreyNet and OpenSIGLE has ensured that preprints, PowerPoint presentations, abstracts, and biographical notes from previous international conferences on Grey Literature are included. OpenSIGLE is thus growing at a remarkable rate. This past year, "700 000 records of the unique European database on grey literature SIGLE migrated to an open access environment" (Giustini, 2010).

SIGLE, the predecessor to OpenSIGLE, assigned nearly 96% of its contents to one of three categories: reports, theses, and conferences. While this classification supports the traditional definition of grey literature along with the most common types of grey material, it can be problematic when creating, analyzing or defining a grey literature typology. For instance, there are several subcategories of reports, ranging from those produced by institutions to annual reports to logs generated by a specific activity. Further, Schöpfel (2006) argues that the theses and conference proceedings distinction fails to take into

account unpublished manuscripts, newsletters, presentations, working papers, preprints, lecture notes, and even personal communications. These documents, regardless of the format they may be presented in, are all types of grey literature, and need to be distinguished as such.

Registry of Open Access Repositories (ROAR)
Founded in 2003, ROAR's key role is providing information concerning the growth and status of open access repositories around the world. Consisting primarily of dissertations and preprints/postprints of peer-reviewed articles, (Digital Library Federation, 2005) ROAR also contains documents in a wide-variety of formats, including multimedia archives. Offering the user an opportunity to present material to the Editorial Review committee for possible inclusion in the repository, ROAR is growing at an exponential rate; as of November 6, 2010, there are a total of 1988 items in the repository, organized into one of 9 repository types (Registry of Open Access Repositories, 2010).

National Repository of Grey Literature (NRGL)
In 2008, a four-year study, *The Digital Library for Grey Literature: Functional Model and Pilot Implementation,* the goal of which is to create a National Repository of Grey Literature (NRGL), was launched. The NRGL originated as an idea in 2005 due to the termination of SIGLE, which greatly affected the state of grey literature in the Czech Republic. Supported by the National Technical Library in Prague (NTK), this project's main goals are the systematic collection, long-term archiving and provision of access to specialized grey literature, especially with regards to research and development, civil service, and education, as well as from the business sphere, marketing "open access" at the national level. To support this goal, the NTK created a network of partner organizations, a functional model, and a pilot application. In addition, on the basis of verified technology and methods defined under the project, recommendations and standards are created for other institutions electing to build their own digital grey literature repositories. These consist primarily of a recommended metadata format, exchangeable designs and templates, examples of licensing models and legal issues, resolved preservation, methodology, archiving, and the provision of access to digital data (National Repository of Grey Literature, 2010). Further details on this project can be found on the NRGL website, http://nrgl.techlib.cz.

Since the end of 2009, a NRGL central user interface has been available to search for grey literature in the Czech Republic. This NRGL central search interface offers a user-friendly system for searching data thanks to data visualization and dynamic contextual navigation. All institutions in the NRGL network are gradually integrated into this interface. At the end of November 2010 there were over 51 000 grey literature records. The interface is available at www.nusl.cz .

Information Register of R&D results (RIV)
RIV is part of the R&D Information System in the Czech Republic. Since 1993, RIV has collected information about the results of R & D long-term intentions and projects supported by different state and other public budgets.

The data available in RIV has been made possible by contributions from public sponsors, namely different ministries and other state offices with the responsibility for a state-run R&D long-term intention and/or R&D project, providing financial aid. This includes the Grant Agency of the Czech Republic, the Academy of Science of the Czech Republic, and local authorities (Research and Development Council, 2006).

RIV refers to the conveyance of data to an informational research system, experimental development, and innovation. According to a disclaimer or "law", the support of experimental development is provided only under the supposition of the truthful publication of pertinent data. As such, RIV contains information about all research results. Nevertheless, RIV does maintain its right to remove information about any results forwarded by experimental institutions, in the event that the data are incorrect, or otherwise deemed inadmissible. In addition, the general terms available for a description of data for RIV include: central evidence of activities, research activity, provider, receiver, other participants, proposal from a research organization, creator, etc.

Register of Publication Activity of the AS CR (ASEP)
The ASEP system is produced under the auspices of the Library of the Academy of Sciences of the Czech Republic. Together with RIV, types of publications are listed according to their form and incorporation. These include monographs, conference contributions, dissertations, electronic documents, conference volumes, temporary publications, articles in a professional periodical, prototypes, norms and rules, specialized maps, certified methods,

software, chapters/sections of books, newspaper articles, patents, reviews, translations, workshops, exhibitions, research reports, and several others. The ASEP system contains bibliographical records concerning research results at institutes of the Academy of Sciences of the Czech Republic from 1985. The ASEP system publication records are also sent to the RIV database (Evidence publikací v AV ČR, 2010). The User interface of the ASEP system is available at www.lib.cas.cz/en/ASEP.

Concluding Thoughts & Future Directions

"Grey is global…grey is growing…grey is good." (Hitson, 2009). Without a doubt, grey literature is here to stay, with the border between grey and white becoming more and more transparent, in response to the increasing number of grey material being posted on the Web. As the above arguments suggest, there is not one set rule of classifying and organizing the grey literature. The typology that has been suggested in this paper is merely one notion of how a typology for this type of material can be implemented; it is certainly not the only one. Nevertheless, subjecting a piece of grey literature that is neither a theses nor a conference paper or report to a 'miscellaneous' or 'other' distinction makes identifying and gathering grey literature that much more challenging.
Implementing a quality control system to guarantee the credibility of grey literature is of vital importance, despite the many challenges.

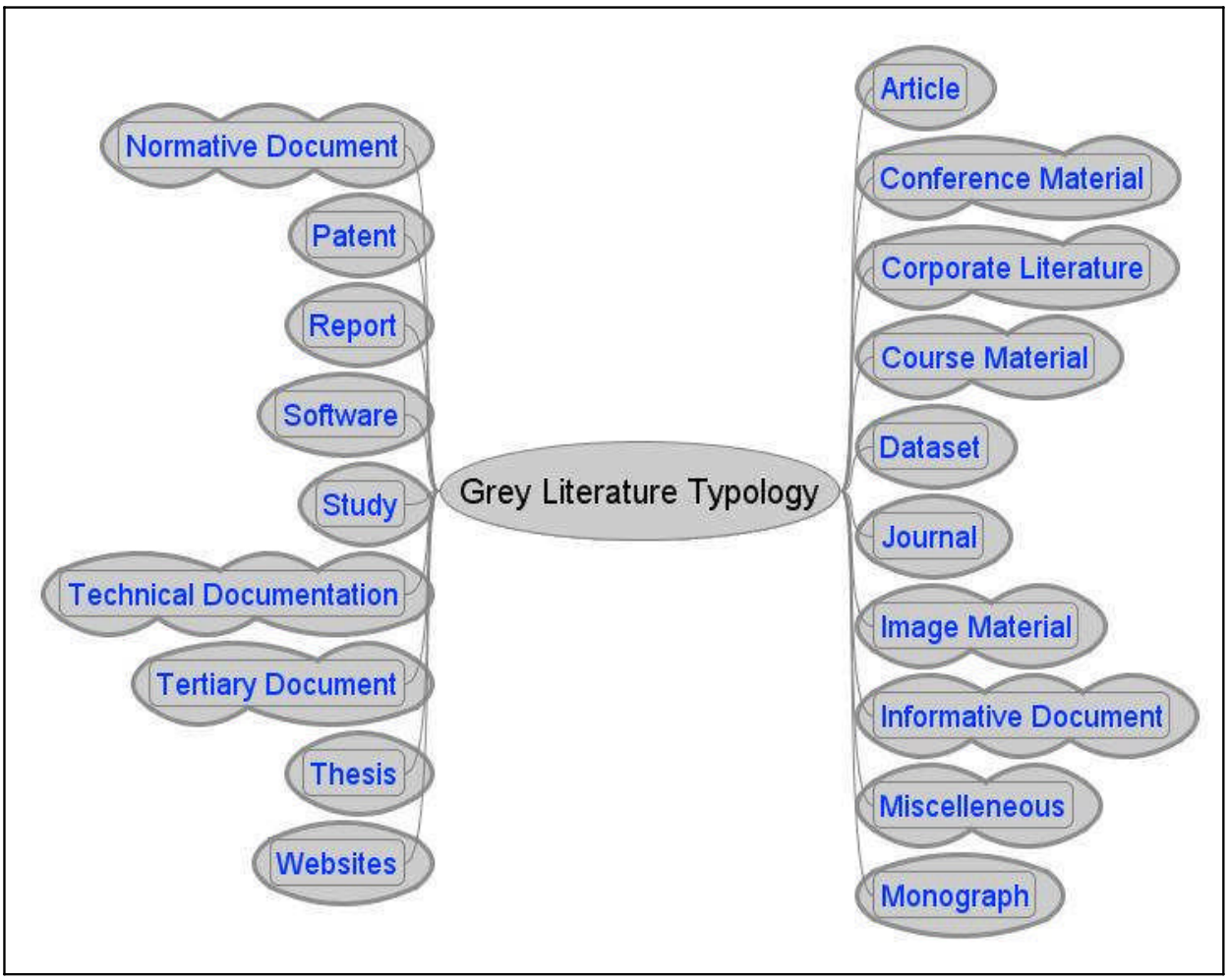
Following this analysis, we established an international Grey Literature Typology Working Group for creating a grey literature typology. The six-member group consists of experts in grey literature, ontological engineering, data modeling, and knowledge organization systems. The Google project platform CODE, a tool for collaborative development, was used for this study. Information about the activities of the working group as well as a link to the grey literature typology project can be found on both the Czech NRGL and GreyNet websites.

A key goal of defining and describing grey literature document types has already been partially addressed; the proposed grey literature typology discussed in this paper has been transformed into open standard machine-readable format as an open available web application, the details of which are available in *Publishing the Vocabulary of the Types of Grey Literature as Linked Data,* a poster presented at the GL12 conference. "The typology of grey literature will be a controlled vocabulary in RDF (Resource Description Framework) expressed as SKOS (Simple Knowledge Organization System) concept scheme. This description of the document types of grey literature has a loose structure with hierarchical relations. Each type will be provided with a definition and a prototype example of a document for which it can be used. By design, it is focused on the description of types. Other documents' attributes, such as content or format annotations, are excluded from the vocabulary." (Grey-literature-typology, 2010)

The first draft of a grey literature typology will be issued in January 2011. The online platform CODE, http://code.google.com/p/grey-literature-typology/, will be open for comments from the grey literature community until the end of March 2011. The Working Group will analyze and respond to comments regarding a clear and structured typology, and compile the materials for incorporation into a draft grey literature typology by the end of May 2011. This recommendation will be included in the first version of a grey literature typology in the SKOS concept scheme, to be published on June 30, 2011. The Development Cycle will be conducted bi-annually, and all versions will be issued via the Web as sustainable linked versions. Future plans and directions will involve the translation of the proposed typology into other languages.

Most in the grey literature community would undoubtedly agree that the use of this material varies in frequency among different disciplines; some subject areas make use of it on a daily basis, while others have not yet skimmed the surface. Nevertheless, grey literature definitely holds a place in the information-seeking behavior of today's researchers, more so than it ever has before. As the future of grey literature information seeking turns towards alternate formats such as multimedia and datasets, the fundamental purpose of grey literature remains the same: increasing awareness of the grey by opening the information doorway and providing unrestricted access to what lies beyond. As Schöpfel (2006) argues, "however diverse, these documents all have one point in common: they contain unique and significant…information that is often never published elsewhere."

Appendix 1: Mind map of the 19 terms identified as narrower term document types in analysis



References

Beissel-Durrant, G. (2004). *A typology of research methods within the social sciences: NCRM working paper.* Retrieved September 12, 2010 from http://www.ncrm.ac.uk/publications/documents/NCRMResearchMethodsTypology.pdf

Burnett, G. (2000). Information exchange in virtual communities: A typology. *Information Research*, *5*(4). Retrieved September 12, 2010 from http://informationr.net/ir/5-4/paper82.html

Di Cesare, R., and Ruggieri, R. (2006). *Evaluation of grey literature using bibliometric indicators: A methodological proposal.* Retrieved September 12, 2010 from https://darchive.mblwhoilibrary.org/bitstream/handle/1912/1322/proc06057.pdf?sequence=1

Digital Library Federation. (2005). *Review of resources.* Retrieved November 6, 2010 from http://www.diglib.org/pubs/dlf106/DLF106part3b.html

Evidence publikací v AV ČR: ASEP (2010). Retrieved November 10, 2010 from http://www.iach.cz/knav/databaze_cz.htm

Giustini, D. (2010). *Got app? Top mobile medicine sites 2010*. Retrieved September 12, 2010 from http://blogs.ubc.ca/dean/2010/04/got-app-top-mobile-medicine-sites-2010/

Grey-literature-typology (2010). *Vocabulary of the types of grey literature*. Retrieved November 20, 2010 from http://code.google.com/p/grey-literature-typology/

GreyNet: Grey Literature Network Service (2010). Retrieved September 12, 2010 from http://www.greynet.org/

Hitson, B.A., and Johnson, L.A. (2009). *WorldWideScience.org: Bringing light to grey*. Retrieved September 12, 2010 from http://opensigle.inist.fr/bitstream/10068/697997/2/GL10,_Hitson_and_Johnson,_Conference_Preprint.pdf

Hjorland, B. (2006). *Document typology*. Retrieved September 12, 2010 from http://www.iva.dk/bh/core%20concepts%20in%20lis/articles%20a-z/document_typology.htm

Library of the Academy of Sciences of the Czech Republic (2007). *ASEP – publication records.* Retrieved January 4, 2011 from http://www.lib.cas.cz/en/ASEP

National Repository of Grey Literature (2010). Retrieved November 9, 2010 from http://nrgl.techlib.cz/

OpenSIGLE: System for Information on Grey Literature in Europe (2010). Retrieved September 12, 2010 from http://opensigle.inist.fr/

Registry of Open Access Repositories, ROAR. (2010). Retrieved November 6, 2010 from http://roar.eprints.org/

Research and Development Council (2006). *RIV* : *Information Register of R&D Results*. Retrieved November 19, 2010 from http://www.vyzkum.cz/FrontClanek.aspx?idsekce=1028

Rose, D. (2010, January 24). Glacier scientist: I knew data hadn't been verified. *Daily Mail Online.* Retrieved November 6, 2010 from http://www.dailymail.co.uk/news/article-1245636/Glacier-scientists-says-knew-data-verified.html

Schöpfel, J. (2006). Observations on the future of grey literature. *The Grey Journal, 2*(2), 67-76.

Schöpfel, J., and Stock, C. (2009). Grey literature in French digital repositories: A survey. Retrieved September 12, 2010 from http://archivesic.ccsd.cnrs.fr/docs/00/37/92/32/PDF/GL10_Schöpfel_Stock_final.pdf

University of North Carolina, Wilmington: Randall Library (n.d.). *Identifying primary, secondary, and tertiary sources*. Retrieved December 31, 2010 from http://www.lib.auburn.edu/bi/typesofsources.htm

https://doi.org/10.26069/greynet-2026-000.555-tgj

A Profile of Italian Working Papers in RePEc [*]

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

Abstract
This paper describes the results of an analysis of Italian Working papers (WP) available both in RePEc (Research Papers in Economics) and in Institutional Repositories (IR) and websites. Given that RePEc is a disciplinary repository based on the active involvement of economic institutions, rather than authors, our analysis intends to explore the institutions' propensity for making their collections available both in disciplinary and Institutional repositories. Therefore, the paper provides a profile of the Italian Economics institutions participating in RePEc as well as an in-depth analysis of the their availability WPs and WP series. Moreover, IRs and websites of the Italian institutions participating in RePEc were analysed to compare the scientific contents available in these important sources of free access information (RePEc, IRs and websites).

Introduction

RePEc (Research Papers in Economic) is one of most important disciplinary repositories, which covers different aspects of research in Economics, and gathers the largest collection of working papers. Founded in 1997, it provides users with a variety of services, ranging from searching facilities for document (IDEAS) as well as research institutions profiles (EDIRC) to a provision of access statistics for items and authors (LogEC) as well as tools for citation analysis (CitEc). This decentralized repository is primarily based on an interconnected network of over 1000 interoperable archives supported by an eclectic mix of participants, from the major commercial publishers, university presses, research centres, central banks to university departments in 80 countries worldwide. This makes RePEc different from other disciplinary repositories. It is based on the collaboration with Economics institutions, which make their collections retrievable by RePEc using both a common bibliographic template for content descriptions and a protocol to exchange data. Only recently has RePEc set up a service that allows single authors to submit their publications, but they are allowed to do it only when they belong to institutions lacking a RePEc archive.

In the framework of the Open access movement, Institutional and disciplinary repositories represent complementary communication channels to enhance the visibility and impact of scientific results. Generally, disciplinary repositories tend to be populated with a greater number of papers compared with the ones available in Institutional repositories (IR). This fact has been underlined in several studies [Swan, 2005; Kingsley 2008; Bjoerk et al., 2010] pointing out that authors show a greater propensity to submit their work to thematic archives rather than self-archiving their works in IRs. Different explanations have been given, ranging from publication practice of specific scientific communities to the capacity of institutions to actively involve the different stakeholders (authors, librarians, information managers, etc.) of the research lifecycle.
Given that RePEc is based on the active involvement of economic institutions, rather than authors, our analysis intends to explore the institutions' propensity for making their collections available both in disciplinary and institutional repositories. For this reason this paper provides first a profile of both Italian economics institutions and their production in terms of Working papers (WP) and WP series listed in RePEc (paragraphs 4.1. – 4.3.) and then analyses whether these institutions make their WP series also available in their IRs and/or web pages (paragraphs 4.4 - 4.5). Therefore, our analysis intends to contribute to the identification of successful strategies to increase the impact of research results within an open access environment.

RePEc characteristics

Even if RePEc is currently listed among the most successful disciplinary repositories (the largest after the well-known arXiv.org), its own founders generally refers to it as a bibliographic service or database [Zimmermann, 2009], a de-centralised non-commercial digital library [Barrueco et al., 2000], a decentralised academic publish system, an open

[*] First published in the GL12 Conference Proceedings, February 2011

library [Krichel, 2001]. This is due principally to two separate, main features: namely its historical development and its organisational model.

RePEc dates back to 1997 and is based on the previous NetEC project founded in 1993 by Thomas Krichel as a collection of projects (BibEc, WoPEc, CodEc, WebEc, BizEc and HoPEc) aiming at distributing information relevant to Economics and in particular focused on WPs diffused via Internet. It is interesting to read about its development in the pre-history of WWW [Krichel, 1997; Karlsson, 1999] because each project represented the effort of both tracking different types of information (from the print working papers in BibEc to the first digital ones in WoPEc, software codes used in Economics, collection of web pages in WebEc, etc.) and progressively establishing the active involvement of different people and institutions in sharing their resources taking advantage of the new information technologies (at the time, gophers, mailing lists etc.).

Moreover the current RePEc is also the result of constructing a cooperation model, which is suited to its scientific community and its publishing practice. For instance the development in 1997 of a centralised "Economic Working Paper Archive" at Washington University based on the model imported from the High Energy community was abandoned in favour of a decentralised database. The economists' "built-in distrust of monopolies" reported by Krichel [2001] coupled with, at that time, over 200 retrievable archives of working papers made more practical to let each institution manage its own collection locally and then make them centrally accessible on a common interface. Therefore, the establishment of a common bibliographic template for content descriptions (called ReDIF) as well as the development of a protocol (called Guildford) to exchange data represented one of the most important achievements, which are still the basic architectural framework of the current RePEc.

Under this perspective RePEc can be considered a distributed digital library, a "collection of metadata records" each one identified by a unique handle, which allows the linkage between records. Hence, the definition of a decentralised bibliographic database, based on the relations between "resource" (i.e. any output of an academic activity: research documents, datasets, computer programs), the resource logical grouping in "collection" (i.e. working paper series and journals), as well as "person" and "institution" [Barrueco, 1999].

These four elements are the core RePEc services, built upon the archives made available by the collaborating institutions on an http or ftp server. They are briefly described hereafter.

- EDIRC (Economics Departments, Institutes and Research Centers in the World) is the service that indexes economics institutions worldwide by countries and fields. It provides detailed information on the institution structure, listing the affiliated "sub-entities" (for instance in the case of universities EDIRC reports the belonging departments, institutes, and/or research centres connected with economics studies). Moreover, EDIRC provides also a list called "*Top 25% Institutions and Economists*" that ranks institutions in each participating country according to a set of criteria described in detail in Zimmermann [2009].
- IDEAS (Internet Documents in Economics Access Service) is the service that provides the user interface to browse and search RePEc scientific contents (journals, working papers, books, book chapters, software components);
- RePEc Author Service allows authors to register in RePEc, proving information on their name variations so that metadata matching as well as work attribution are facilitated. Authors also provide information on their affiliations. Only when registered is an author ranked together with his/her institution and obtains notifications of new citations found in RePEc.
- MPRA (Munich Personal RePEc Archive) is the only central archive where researchers can submit their works only when their affiliated institution does not participate in RePEc.

Besides organising this collection of archives and making their data freely available, RePEc provides a set of additional intermediary user services:

- NEP (New Economics Papers) is a notification service of new downloadable WPs for over 40 specific fields. Voluntary editors compile subject specific reports that filter RePEc new additions to provide subscribers with update information constituting a "simple form of peer review" [Bátiz-Lazo, 2005].
- LogEC is a service that provides access statistics for each item.

- CitEC (Citation in Economics) is the service that provides an autonomous citation index of many electronic documents distributed by RePEc (74% at the time of writing). This service is maintained by Josè Manuel Barrueco at the University of Valencia.

The variety of services provided by RePEc coupled with different functionalities and information needs succeed to "describe the discipline, rather than simply the documents" [Krichel, 2001] produced by the Economics scientific community.

Objectives and Methods

Our analysis was driven by the type of organisation model that characterises the RePEc interconnected network and in particular by the active role played by Economics institutions in making their collection retrievable, thus taking advantage of RePEc services to increase their visibility and impact within an international scientific community. Under this perspective RePEc is worth analysing for the following main reasons: a) unlike the majority of e-print archives, it is based on collaboration with different Economics institutions which make their collections retrievable by RePEc, or more precisely by the IDEAS Service, b) it is the largest open source of WPs in Economics and related fields, and c) WPs contribute to determine institutions' and authors' ranking positions measured by RePEc bibliometric analysis. For these reasons RePEc as well as the IRs and websites maintained by the Italian institutions participating in RePEC are valid sources of analysis to achieve our overall objective, that is to explore the institutions' propensity to make their collections available both in disciplinary and institutional repositories in Italy.

Our analysis is divided into two parts. In the first we have analysed the Italian contribution to RePEc in order to identify, on the one hand types of economics institutions participating in RePEc, and on the other, features and characteristics of the WPs made available in this disciplinary repository. This analysis was performed using the following RePEc data sources:

a) EDIRC list of entity and sub-entities participating in RePEc;
b) IDEAS list of Italian WP series;

The analysis of both Italian institutions participating in RePEc and their WP collections was performed at two levels. In the first data was gathered from the entire set of WP series listed in IDEAS, while in the second, data was collected from the WP series reported in the list of "*Top 25% Institutions and Economists in Italy".* This list ranks institutions on the basis of the number of authors registered in RePEc Author service, Institutions listed in EDIRC, bibliographic data collected by RePEc, as well as citations counted by CitEc and access statistics registered by LogEc. The comparison of the data of the entire set of Italian contribution with that contained in the list of the best-ranked institutions allowed us to identify the share production of the best-ranked Italian institutions, in terms of consistency, stability over time, etc.

In particular, the Italian WPs contribution in RePEc was analysed in terms of:

- WP series characteristics, considering:
  - Number of series registered;
  - Longevity (live, dead series),
  - Vitality (young and new-born series);
- WP characteristics, considering:
  - Number of WP for each series;
  - Average number of WPs per series and per year.

Within the Italian WP series we excluded those produced by and/or in collaboration with commercial publishers that turned out to be journal collections.

As RePEc is continuously updated, the analysis of Italian WP series referred to the month of July 2010.

In the second part of our study the best-ranked institutions listed in the *Top 25%* were considered in order to determine whether:

- These Institutions had an IRs;
- The series listed in IDEAS were available also in the IRs and/or in websites;
- The documents' temporal coverage in IRs and/or websites corresponded to that registered in RePEc;
- IRs and/or websites provided access to other types of GL documents.

This analysis also included those institutions that are listed in the *Top 25%*, but do not register any WP series in RePEc.

Results

Italian WPs in RePEc in a European context

The production of more than 80 countries worldwide is retrievable in RePEc. To give an overview of the Italian production of WPs at international level, we selected the WPs made available by United Kingdom, Germany, France and Spain. These countries were selected because they were among the major contributors to the SIGLE database.

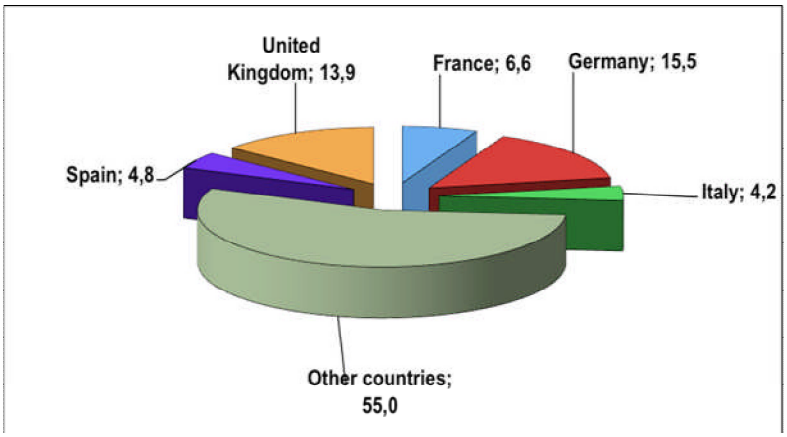


Figure 1 . - WPs of some European countries in RePEc (%)

Italy participates in RePEc with 15064 WPs, which is 4% of the entire database. The major WP producers in our sample are: Germany with 55062 WPs and the United Kingdom with 49425 WPs. The French and Spanish contributions are closer to that of Italy.

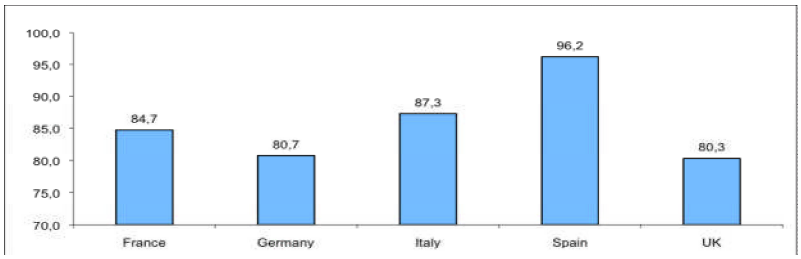


Figure 2. – Online WPs of some European countries in RePEc (%)

If we consider the online WPs (fig. 2) we can notice that most of them are available in full text, with Spain reaching the highest percentage, followed by Italy and France.

Italian institutions in RePEc

Table 1 shows the composition of the Italian participation in RePEc. Within a total number of 369 providers, the major contributors are universities represented by their departments, faculties, and centres, reaching all together 71.2%. Other types of organisations are Research centres, Foundations, Governmental institutions and Associations.

Table 1. – Number and percentage of Italian institutions in REPEC and in Top 25% by type of providers

Provider	No.	%	Top 25% institutions No.	Top 25% institutions %
University department	130	35.2	45	51.1
University centre	82	22.2	10	11.4
University faculty	51	13.8	20	22.7
Research centre	31	8.4	8	9.1
Foundation	22	6.0	4	4.5
International organization	8	2.2	1	1.1
Governmental institution	13	3.5	--	--
Association & Society	32	8.7	--	--
Total	**369**	**100.0**	**88**	**100.0**

Considering the composition of the Italian providers listed in the *Top 25%* (table 1), there are 88 best-ranked institutions; the majority of them are University departments, followed by faculties and University centres. Compared with the figures of the total number of Italian institutions, we can notice that the presence of University sub-entities is more relevant increasing in particular in the case of University departments (51.1% vs. 35.2%), Faculties (22.7% vs. 13.8%) as well as for Research centres (9.1% vs. 8.4%).

Italian WP collections in RePEc

Working papers have long been a mainstay of scientific output in Economics. Like other types of GL documents, WPs attenuate the gap of publication delay and editorial space limits, thus proving updated and comprehensive information on specific research areas. As reported in Krichel [2001], "Economists do not issue preprints as individuals; rather economic departments and research organizations issue working papers". Additionally it is important to point out that Economics institutions generally organize this type of report literature in collections, providing for each of them a title that specifies the topic of the series and progressive numbers in the collection (known as report number). This editorial activity differs from the one carried out by commercial publishers in two aspects: there is no predefined number of issues/WPs to be published in one year, and generally there is no formal peer-review procedure that is generally dependent on internal institutional rules and practices. Nevertheless, many of these WPs represent important reference points in the scientific literature: they are often cited in journals and may represent the official position of important institutions in key issues. Hence, the importance of their free, rapid and wide diffusion through different channels, Institutional repositories, websites and of course disciplinary repositories like RePEc that allows users to gain a more comprehensive picture of the production of WPs at international level.

Given these premises, our hypothesis is that the editorial activity carried out within an Economics institution in issuing and managing WP collections is comparable with the commercial publishing process of scientific journals. Therefore, in the measurement of the consistency of WPs and WP series we focused our attention in particular on:

- The *stability* of collections, i.e. their continuity over time, that according to our hypothesis indicates a well consolidated research area, whose results are progressively issued and circulated under a specific series title;
- The *novelty* of collections, i.e. the issue of new collections that indicates the setting up of new research areas in which institutions are acquiring scientific results that need to be diffused, generating an *ad hoc* new series. That is a similar procedure to that which is adopted when new journals are launched.

To describe these characteristics we used two indicators: *longevity* and *vitality*.

To measure "longevity", we classified WP series as following:

- *Live WP series*, that is series that are still available in RePEc for the years 2009 or 2010;
- *Dead WP series*, that is series that are no more available either for the years 2009 or 2010.
- To measure "vitality", within the live series, we classified WP series as following:
- *Young series*, that is series available in RePEc from 2007 on;
- *New born series*, that is series available in RePEc since 2009 or 2010.

Moreover, considering that each provider produces different numbers of WPs over different periods of time, we measured the WP consistency in terms of average number of WPs within each series and the average annual contribution, as following:

- *Series' average weight*: is the average number of WPs contained in a series;
- *Annual average contribution:* is the average number of series provided by an institution in one year.

These indicators have been applied to the entire set of Italian WP series as well as to the series produced by the best-ranked institutions.

Characteristics of WPs series: "longevity"

There are 145 WP series listed in RePEc (table 2), the major producers are University departments (53.8%), Research centres (14.5%) and International organisations (15.9%). According to the previously provided classification of *longevity*, the majority of RePEc providers contribute with series that have a stable and continuous production over the years (109 out of 145, equal to 75.2 are *live* series), and this is particularly evident in the case of Foundations, Governmental institutions, University departments, and Research centres. The only exception is represented by International organisations that have a high percentage of series that finish in 2008 or before. These collections are produced by FAO and UNICEF, the latter in particular has a high number of *dead series* that contain project results. Therefore we can assume that a collection is closed when the project is finished, following the project lifecycle.

Table 2. - RePEc series according to "longevity"

Provider	RePEc series					
			Live		*Dead*	
	No.	%	No.	%	No.	%
University department	78	53.8	64	82.1	14	17.9
University centre	13	9.0	11	84.6	2	15.4
University faculty	3	2.1	2	66.7	1	33.3
Research centre	21	14.5	18	85.7	3	14.3
Foundation	4	2.8	4	100.0	--	--
International organization	23	15.9	8	34.8	15	65.2
Governmental institution	1	0.7	1	100.0	--	--
Association & Society	2	1.4	1	50.0	1	50.0
Total	**145**	**100.0**	**109**	**75.2**	**36**	**24.8**

Table 3. – Best-ranked RePEc series according to "longevity"

Provider	RePEc series					
			Live		*Dead*	
	No.	%	No.	%	No.	%
University department	41	58.6	37	90.2	4	9.8
University centre	9	12.9	9	100.0	--	--
University faculty	--	--	--	--	--	--
Research centre	14	20.0	12	85.7	3	14.3
Foundation	4	5.7	4	100.0	--	--
International organization	2	2.9	1	50.0	1	50.0
Total	**70**	**100.0**	**63**	**90.0**	**7**	**10.0**

The analysis of the longevity indicator applied to WP series produced by the best-ranked institutions is reported in table 3. Comparing these data with the entire set of WP series in RePEc, we can notice that the percentage of live series increases reaching 90% in total providing a higher degree of stability and continuity in production of the best-ranked institutions.

Moreover, it is worth underlining that 70 WP series are produced by the best-ranked institutions representing almost 50% of the total number of the Italian WP series listed in RePEc and this confirms that WP production contribute to determine the positions of the institutions' ranking.

Characteristics of WPs series: "vitality"
The analysis of the vitality indicator applied to the entire set of Italian RePEc WP series is reported in table 4.

Table 4. - RePEc series according to "vitality"

Provider	Live RePEc series					
			Young		*New-born*	
	No.	%	No.	%	No.	%
University department	64	58.7	8	7.3	3	2.8
University centre	11	10.1	1	0.9	1	0.9
University faculty	2	1.8	--	--	--	--
Research centre	18	16.5	1	0.9	4	3.7
Foundation	4	3.7	--	--	--	--
International organization	8	7.3	1	0.9	--	--
Governmental institution	1	0.7	--	--	--	--
Association & Society	1	0.9	--	--	--	--
Total	**109**	**100.0**	**11**	**10.1**	**8**	**7.3**

Among the live series, there is a low percentage of young series (10.1%), which are mostly produced by University departments, while new-born series (7.3%) are in particular produced by Research centres (3.7%). This highlights that such production is concentrated within research environments.

Table 5. – Best-ranked RePEc series according to "vitality"

Provider	Live RePEc series					
			Young		*New-born*	
	No.	%	No.	%	No.	%
University department	37	58.7	4	6.3	1	1.6
University centre	9	14.3	2	3.2	--	--
University faculty	--	--	--	--	--	--
Research centre	12	19.0	1	1.6	3	4.8
Foundation	4	6.3	--	--	--	--
International organization	1	1.6	--	--	--	--
Total	**63**	**100.0**	**7**	**11.1**	**4**	**6.3**

The same classification applied to the series produced by the best-ranked institutions (table 5) gives similar results. There is a small increase in the percentage of *young* WP series, while *new born* have similar values compared to the results obtained in the entire set of Italian WP series in RePEc.

Characteristics of WPs series: WPs production over time

As seen in paragraphs 4.3.1. and 4.3.2., Italian WP series in RePEc represent stable collections, which provide scientific results on institutional consolidated research areas over a continuous period of time. Now, in order to verify whether WP collections produced by Italian Economics institutions in RePEc also have features similar to scientific journals, we analysed the number of WPs within each series and measured the average number of WPs within each series as well as the average annual contribution.

Table 6. – RePEc annual Italian contribution by type of provider

Provider	No. of series	No. of WP	Series' average weight	Annual average contribution
University department	77	6671	86.6	11.7
University centre	14	1542	110.1	15.0
University faculty	3	140	46.7	4.2
Research centre	21	1651	78.6	11.0
Foundation	4	1380	345.0	37.3
International organization	23	614	26.7	4.5
Governmental institution	1	18	18.0	18.0
Association & Society	2	73	36.5	9.1
Total	**145**	**12089**	**83.4**	**11.6**

Table 6 shows the results of this analysis. 145 series contain a total number of 12089 WPs. The general average number of WPs is 83 papers for each series, while the average of WPs produced in each year for each series is 11,6. Considering now the different types of providers, Foundations produce on average a very high number of WPs, followed by University centres and University departments. We encountered a similar result when we measured the annual average contribution.

Table 7. - RePEc annual Italian contribution by type of best-ranked provider

Provider	No. of series	No. of WP	Series' average weight	Annual average contribution
University department	41	4573	111.5	12.5
University centre	9	1305	145.0	15.5
University faculty	--	--	--	--
Research centre	14	1496	106.9	14.4
Foundation	4	1380	345.0	37.3
International organization	2	58	29.0	5.3
Total	**70**	**8812**	**125.9**	**14.6**

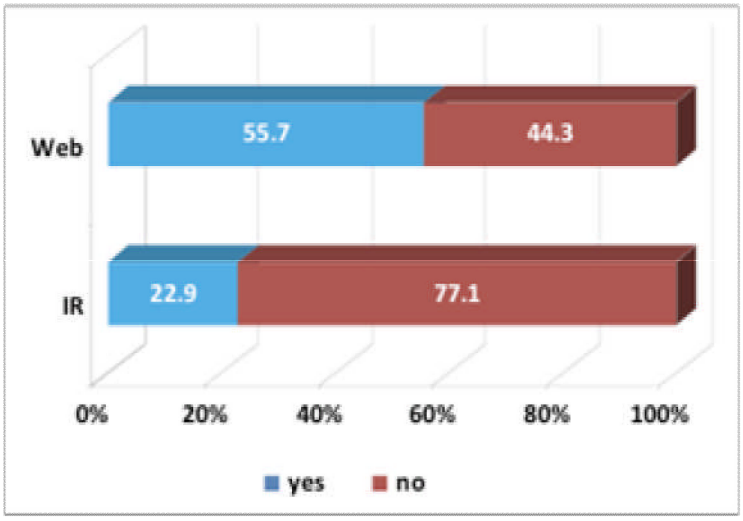
The same measure applied to the top ranked institutions (table 7) shows first of all that the general average is much higher, with a higher number of WPs for each series (125.9), confirming that Foundations and University centres produce a number of WPs above the average. Further, the average production increases to 14 WPs per year; University centres, Research centres (15.5) and Foundations (37.3) produce an above general average number of WPs.
Given this data we can compare WPs series in Economics with commercial journals considering that journals produce at most 12 issues per year and that the number of articles in one journal issue is generally much lower that the number of WPs produced in one series.

Moreover, we can notice that 73% of WPs (i.e. 8812 out of 12089 total number of WPs) is produced by the best-ranked institutions and this confirms once again that this production contributes to determining their position in the ranking, being part of evaluation metrics used in RePEc.

Italian Institutions in RePEc and their IRs and websites

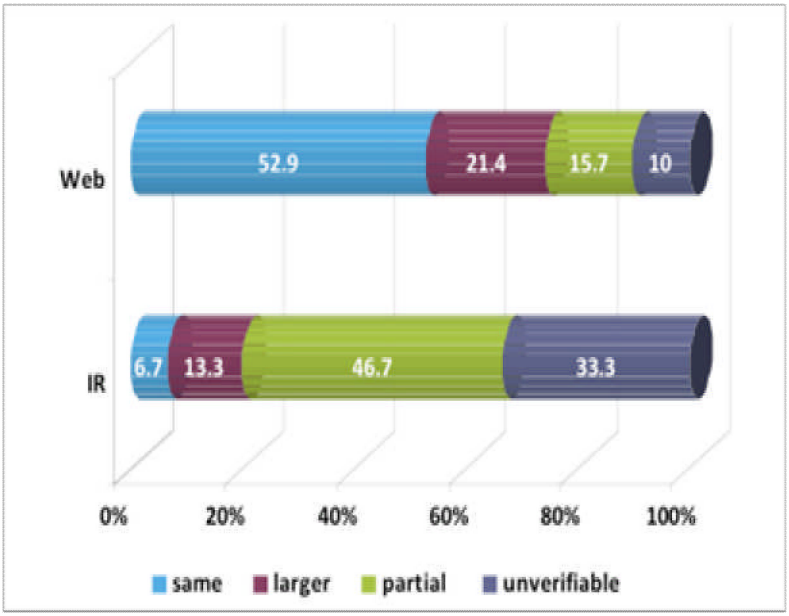
In this part of the study we analysed whether the Institutions participating in RePEc also populate their own IRs and/or diffuse their WP series within their websites. For this reason, sources of our analyses were the IRs of the 88 Italian top ranked institutions as well as their websites.

Fig. 3. Availability of RePEc series in local IRs and websites



Considering University departments, centres and faculties belonging to the same university, there are 35 RePEc providers that can potentially submit their scientific production in the 17 local IRs. Only 22.9% (equal to 8 providers) makes RePEc series also available in their local IRs. Analysing the websites of the 88 providers listed in the *Top 25%*, the situation is different. 55.7% of them (equal to 49 out of 88 providers) list the RePEc WP series in their websites too (fig. 3).

Fig. 4. – Temporal coverage of RePEc WP series in local IRs and Websites



A more detailed analysis was performed on the WPs series available in local IRs and websites in order to verify if there is uniform information content in RePEc, local IRs and websites. To do so, we checked the temporal coverage of each series within local IRs and websites (fig.4). Considering Institutional repositories, 6.7% of series produced by the best-ranked institutions in RePEc have the same temporal coverage compared with that available in RePEc, 13.3% a larger temporal coverage, while the majority of these institutions (46.7%) provide a limited temporal range of WP series in their IRs.
For 33.3% of them this data could not be retrieved, and this depends on how repositories organise the bibliographical data of their collections as well as on the ways the scientific community (i.e. University departments, centres and/or scientific groups) are associated with the collections they produce.
Considering websites, 52.9% of WP series have the same temporal coverage as in RePEc, while 21.4% make their WP series available for a longer period. The differences in temporal coverage can be interpreted in different ways. The richness of information provided in websites can be explained by the fact that research teams generally directly manage their own websites and are obviously directly interested in providing an updated and comprehensive picture of their research activities and results. IRs are unfortunately often managed "outside the scientific community" and further efforts in building efficient collaboration among the different stakeholders (researchers, librarians, information managers) still need to be made. Similar results were also obtained in [Bjoerk et al., 2010].

Availability of other GL information in local IRs and websites

To conclude our analysis we also checked both IRs and websites to identify whether they contained other types of GL documents. Our intention was to verify if the information providers select types of documents and/or scientific content to be made available locally. We found that both IRs and websites provide information on other GL documents (fig. 5). In the case of IRs (42.9%), they are generally course materials and theses, while in 55.7% of websites they are mainly proceedings, workshops and data set.
If we consider other WPs series available both in IRs and websites, we can see that there is no evidence of information providers selecting the WPs to be retrievable on their local systems, only 11.4% provide information on other WPs series in IRs and 22.9% in websites.

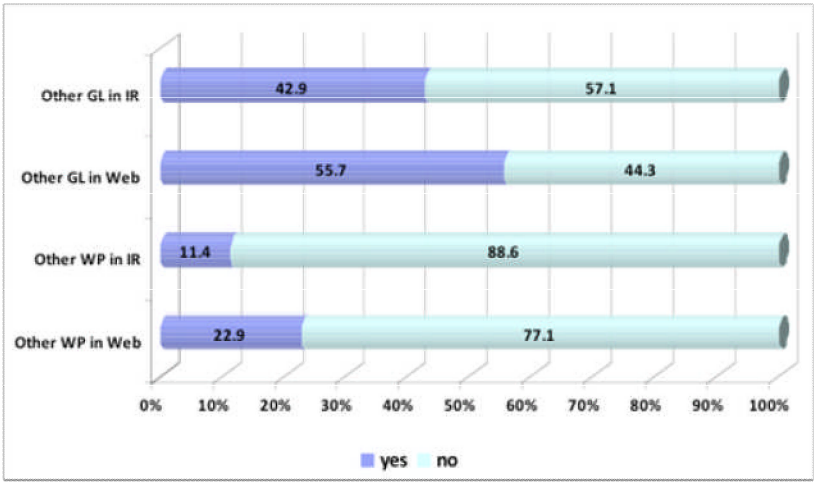


Fig. 5. – Availability of other GL documents in IRs and websites

Conclusions

This study provided a profile of the Italian Economics institutions participating in RePEc associated with the analysis of their production of WPs and WP series. The results of our analysis indicate that among the 369 institutions participating in RePEc the major contributors are academies represented by their belonging to University departments, faculties and centres. They also constitute the core of the institutions listed in the *Top 25%* together with Research centres and Foundations. The analysis of WP series made available in RePEc confirms the important role played by this type of GL document. They contribute to determining the ranking position of the institutions (75% of the total number of WPs are produced by the best-ranked Italian institutions) constituting a large part of the free documents, on which RePEc bibliometric services measure institutions' and authors' impact, using various criteria to determine citations and access statistics [Zimmermann 2009].

Moreover, they represent the main research areas covered by scientific institutions whose results are diffused in a continuous and stable way within WP collections: the majority of WP series are *live*, current collections. Our hypothesis that the institutional activity of editing Economics WP series has features similar to the publication of journals was confirmed by the analysis of WPs contained in the series. In fact, measuring the average number of WPs per series and per year, we found that they contain on average a high number of WPs as well as figures comparable with the publication of issues within scientific journals. This figure is even higher if we consider the WPs produced by the best-ranked Italian institutions. They publish on average more than 14 WPs per year within a series, with peaks above the average in the cases of University centres, Research centres and Foundation.

The results obtained in the second part of the study that aims to analyse the availability of WPs in both Institutional repositories and websites maintained by the 88 best-ranked institutions in RePEc provided a more fragmented picture. In fact, IRs do not seem to represent a preferential channel for WP diffusion: a limited number of Italian institutions additionally makes RePEc series available in their local repositories, the temporal coverage of these collections in IRs is in most cases briefer than in RePEc. Moreover, the identification of RePEc series associated with their producers (University departments, centres or research groups) within IRs is not always a straightforward procedure, depending on the way IRs organise their scientific contents. These aspects may bias the retrieval of this important source of information as well as the visibility of their direct producers.

On the contrary, the websites maintained by the 88 top ranked institutions reflect their scientific production well. In fact, there is a more complete correspondence of WPs listed in RePEc, both in terms of collections available and temporal coverage. Of course the direct involvement of research teams in the development of websites contribute to the achievement of a rich and comprehensive description of their scientific activities and products.

To conclude, our study has outlined a contradictory picture of the propensity to diffuse scientific contents in both disciplinary and local IRs. On the one hand, Italian institutions voluntarily contribute with their WPs and WP collections to build in RePEc a critical mass of information relevant to Economics, on the other, they seem to neglect the role of local IRs in diffusing their scientific results. However, the richness and comprehensiveness of the information available in websites supports the hypothesis that in Italy IRs have not yet succeeded in achieving an active and efficient collaboration among the main stakeholders of the research lifecycle (researchers, librarians, information managers, IR's developers). Of course, this hypothesis has to be confirmed carrying out future, *ad hoc* analysis that can contribute to foster Open access also at local levels.

References

Barrueco Cruz Josè Manuel, Krichel Thomas, (1999). "Cataloging economics preprints: an introduction to the RePEc project", *RePEc and ReDIf documentation*, Shankari, RePEc Team, available at: http://openlib.org/home/krichel/papers/shankari.html

Barrueco Cruz Josè Manuel, Krichel Thomas, (2000). "Distributed cataloging on the Internet: the RePEc project". In: *Metadata and organizing educational resources on the internet*. Haworth Information Press. pp.227-241, available at: http://eprints.rclis.org/bitstream/10760/4010/1/jagt.html

Barrueco Cruz Josè Manuel, Krichel Thomas, (2005). "Building an autonomous citation index for grey literature: the economics working papers case". *The grey journal, an international journal on grey literature*, 1 (2).

Batiz-Lazo Bernardo, Krichel Thomas, (2005). "On-line distribution of working papers through NEP: A Brief Business History". *EconWPA*, available at: http://129.3.20.41/eps/eh/papers/0505/0505002.pdf

Björk B-C, Welling P, Laakso M, Majlender P, Hedlund T, et al. (2010). "Open access to the scientific Journal Literature: Situation 2009". *PLoS ONE* 5(6), available at: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0011273

Foster Nancy Fried, Gibbons Susan, (2005). "Understanding faculty to improve content recruitment for Institutional Repositories". *D-Lib Magazine*, 11 (1), available at: http://www.dlib.org/dlib/january05/foster/01foster.html

Karlsson Sune, Krichel Thomas, (1999). "RePEc and S-WoPEc: Internet access to electronic preprints in Economics". *RePEc and ReDIf documentation*, RePEc Team, available at: http://openlib.org/home/krichel/papers/lindi.html

Kingsley Danny, (2008). "Repositories, research and reporting: the conflict between institutional and disciplinary needs". Paper presented at the VALA 2008: *Libraries/changing spaces, virtual places*, available at:

http://dspace-prod1.anu.edu.au/bitstream/1885/46096/1/117_Kingsley_Final.pdf

Krichel Thomas, (1997). "About NetEC, with special reference to WoPEc". *Computers in higher education economics review,* 11 (1), available at: http://www.economicsnetwork.ac.uk/cheer/ch11_1/ch11_1p19.htm

Krichel Thomas, (2001). "RePEc, an open library for Economics". In: *The Economics and usage of digital library collections*, Ann Arbor, Michigan (US), Revised version 2001 available at:

http://eprints.rclis.org/bitstream/10760/12154/1/salisbury.a4.pdf

Krichel Thomas, Zimmermann Christian (2009). "The economics of open bibliographic data provision". *Economic analysis and policy*, 39 (1), available at: http://ideas.repec.org/a/eap/articl/v39y2009i1p143-152.html

Swan Alma, Brown Sheridan, (2005). "Open access self-archiving: An author study". [Departmental technical report], available at: http://cogprints.org/4385/1/jisc2.pdf

Zimmermann Christian, (2009). "Academic rankings with RePEc". Working papers 2007-36, University of Connecticut, Department of Economics, revised Mar 2009, available at: http://www.econ.uconn.edu/working/2007-36r.pdf

https://doi.org/10.26069/greynet-2026-000.555-tgj

https://doi.org/10.26069/greynet-2026-000.555-tgj

From OpenSIGLE to OpenGrey: Changes and Continuity *

Christiane Stock and Nathalie Henrot *(France)*

Abstract
First presented at the GL8[1] conference in New Orleans 2006 as a prototype, OpenSIGLE went life in December 2007. After three years of existence, the results are beyond all expectations. OpenSIGLE has become a reference source for grey literature, and its user community has grown constantly, especially from outside Europe. The integration of the GL conference preprints into the repository from 2008 onwards not only added research papers on the topic of grey literature to its contents, but also permitted OpenSIGLE to be accepted in the "Directory of Open Access Repositories" (OpenDOAR).*
In spite of the success of OpenSIGLE it's not wise to rest on one's laurels. The change of name to "OpenGrey" signifies a shift in the content of the repository as well as in its physical appearance. Besides providing a new look and a more convenient technological environment, OpenGrey closes the gap between the close of the SIGLE database and today, including recent records and links to the full text.*
The paper presents the new website which includes numerous facilities requested by users such as OAI-PMH, the possibility to export records and an improved access to the document itself. OpenGrey also takes into account a changed user behaviour, where visitors arrive after searching Google or Google Scholar* and want all relevant information at a glance. The paper further explains input procedures and gives other information for the ongoing updates of the repository. Finally we call former SIGLE members and new partners to contribute to OpenGrey.*

From SIGLE to OpenSIGLE

SIGLE (System for Information on Grey Literature in Europe) was a unique European database of bibliographic records in grey literature. It was produced between 1980 and 2005 by initially seven and in the end fifteen members of the European Union, represented by major libraries and research organizations. Its contents covered all scientific disciplines (pure and applied science and technology, economics, social sciences and humanities).
As a commercial product SIGLE was accessible through subscription to hosts, e.g. STN International, and available on a CD-ROM produced by Silverplatter/Ovid.

INIST decided to transfer the results of 25 years of work[2] onto an open access platform. As a result OpenSIGLE went live in December 2007 with almost 700 000 bibliographic records, using DSpace* technology and its qualified Dublin Core metadata* format. The operation was to be low cost and easily feasible.

It was very important that key information found in every SIGLE record was preserved during the migration:
- English title or keywords
- SIGLE classification code
- Availability statement, in order to facilitate the order of a paper copy

One of the main goals of SIGLE was to facilitate access to the paper document. Most of today's requests for assistance from users concern the document availability.

OpenSIGLE – its evolution and its usage

Though not perfect[3], with its lack of flexibility for the user, and no possibility of exporting either search results or records, OpenSIGLE found an ever growing audience and new visitors from – nearly - all around the world.

OpenSIGLE was included in OpenDOAR (Directory of Open Access Repositories) in November 2009, after its integration in the WorldWideScience.org* portal a year earlier. Indexed in Google and Google Scholar since summer 2008, the database is reached through the Google search engine by an increasing number of users, and visits via Google Scholar amount to 30% per month. For 2010 the average number of monthly visits exceeds 35 000. The audience is definitely worldwide, not only European: 10 % of the visitors come from Asia, 20% from North America, continents not involved in the database production. These proportions have remained fairly stable over the past year.

* First published in the GL12 Conference Proceedings, February 2011

The following statistics (see below) are based on php/MyVisites*, an open source software using a tracker, similar to Google analytics. As a result, figures are lower than with a log analyser[4].
Two years of statistical data collection reveal a considerable evolution in numbers and the influence of the Google generation. Between 2009 and 2010 the number of visits almost tripled, and the number of page views more than doubled. Simultaneously the average duration of a visit decreased as well as the number of pages viewed in a visit.

We take the data as a sure indication of a change of behaviour in our visitors. Many of them reach the OpenSIGLE after using a search engine, spend a short time looking for relevant information and leave. However some of our users tried to replicate former search strategies in OpenSIGLE and were successful.

Average per month	2009	2010
Visits	12 000	35 000
Pages viewed	35 450	88 580
Length of visit	107 sec	91 sec
Number of pages viewed per visit	3.4	2.8

The most important highlight for 2010 relates to the Greynet collection[5]. As of October 2010 the complete collection of conference preprints from the GL conferences are accessible in full text and in open access. GL5 to GL11 were provided by GreyNet in electronic format and the first conference became available in full text on OpenSIGLE as early as May 2008. GL1 to GL4 were received in paper form, digitized by INIST and uploaded to OpenSIGLE in September 2009.

When OpenSIGLE became OAI-PMH compliant, we decided to be more explicit about the rights and adopted the Creative Commons licence*, in order to inform the users and partners about the different conditions of use of the bibliographic data and GL proceedings.

Time for change
Several reasons lead us to think of changing the system.
Right from the beginning of OpenSIGLE we were facing limits in the technical performance of DSpace, especially when uploading and indexing GL collections. The problems were even noticeable for smaller updates. DSpace obviously reached its limits with 700 000 records in the database. Besides, the website layout was rather simple (low key) and today no longer answers current needs for referencing.
The most important reason for thinking of a change was the absence of features requested over the years by OpenSIGLE users, such as the possibility to export search results for state of the art studies.

An important number of emails from users request how to obtain the document, a service which was one of the principal goals of SIGLE and OpenSIGLE. This essential information should be seen at first glance.
In fact, the availability statement is « hidden » in the full record display of OpenSIGLE. However, few users take the time to look it up. 55 percent of the visits only display one page in OpenSIGLE before leaving again, according to our statistics. In addition, in the context of the Google generation we presume that the brief record display becomes more and more the "entry page" to the database, due to the Google Scholar search, and half of the time the unique page-view. The new website should therefore improve the access to information on document supply through the general layout, improved record display and additional links.

Early on, INIST-CNRS intended to re-open OpenSIGLE for "new" input. This can be done in two ways:
- add records from 2005 onwards from former EAGLE* members (INIST-CNRS included).
- open the database to new European partners.
However, the technical limits mentioned above would have been a major hindrance, hence the need for a change of software.

In order to reinforce the changes in the technical environment as well as in the contents and policy for the database, INIST-CNRS decided on a change of name: "OpenGrey". Several new domain names have been acquired for OpenGrey with the extensions .eu, .fr, .net and .org. The new domain name will be *opengrey.eu*.

The OpenGrey homepage
The homepage of the new website OpenGrey aims to meet current needs of the users as well as to facilitate referencing. It is divided into three parts with different groups of information. The upper field holds the name and the logo, - inspired by the lemniscates or infinity symbol -, and provides information on the contents for referencing. Tabs allow the user to access further information, e.g. on the partners and former EAGLE members as well as to choose the language for the user interface.
Besides the "Google" like search field, the centre includes three blocs for short texts: a mini "about", a "focus" on specific subjects and a "news" bloc.
The bottom part, separated into four blocs, provides more information on search and help, a choice of export tools (highly requested by users), legal mentions, and further information (about, tools for partners etc.).

Change and continuity
The major change from OpenSIGLE to OpenGrey relates to the software: The DSpace platform used for OpenSIGLE is replaced by Exalead® * as the search engine used for the database, completed with in-house developments using php and MySql software for the user interface.

Persistent identifiers are essential to guarantee perennial access to records or documents. In OpenSIGLE each record as well as the communities and collections of the DSpace architecture are identified by a unique identifier, the handle*. The handle system allows to using the URLs as persistent identifiers which remain, even in case of server changes. Many websites linking to OpenSIGLE or a particular collection (e.g. GL11) use the handle. In order to assure continuity the handles of the present OpenSIGLE records will migrate to the new system. A redirection is planned for the handles identifying communities and collections of OpenSIGLE which cease to exist in the same way in OpenGrey.

Another change comes with the Exalead software. The user interface for search provides the Google-like field for full text record search. Exalead also offers the possibility to refine the results through faceted search. The criteria chosen for refinements are similar to the Czech repository NUSL/NRGL[6] (author, subject, date, language…). This feature has a major consequence for the backoffice: metadata which were merged for the simpler DSpace format in OpenSIGLE need to be detailed once more, and even new controlled fields must be added to allow the refinements and to insure consistency.

As mentioned before, we know from usage statistics that the brief record display is the entry page for many visitors and the one and only page viewed by half of them. Therefore the development of features based on this record display has become a necessity. All relevant information must be available at first glance or invite the user to click further. For example, we intend to make an explicit distinction between:

Access to the paper copy
Access to the online full text

In addition we must provide easy links to both fields and to further information, e.g. on how to obtain a copy of the paper document or on the terms and conditions of the partner organization.

The partner homepage
OpenGrey will provide each partner with a space for information on its institution, a kind of "homepage", accessible from a tab "Partners" on the OpenGrey homepage. OpenSIGLE already offers information on former EAGLE members in the collection page of the "Country

Community"[7]. This content will be transferred to the new partner homepage as a starting point, namely:
- General information of the organization, with a logo, if available.
- Information on how to obtain a copy from old SIGLE records.

Each partner is strongly encouraged to propose any improvement or information about its organization, its policy of document delivery and the fees attached etc… Any information on national grey literature initiatives, databases, recent updates, or any information on conferences relating to the subject could be announced in the "News" section of the Homepage.

Contents, input and updates of OpenGrey

OpenGrey will include all records from the present OpenSIGLE database. In addition, we intend to add new records (created since 2005) from partner organizations, including French grey literature as mentioned before. OpenGrey will not host the documents themselves. The new cooperation will be based on formal agreements with each partner organization. The partner retains the control over the grey documents. An open access policy is encouraged for the full text documents, items which remain under the control of the partner. The bibliographic records as well as the OpenGrey website will be placed under the Creative Commons Licence.

OpenGrey metadata should be as rich as possible. This is why we prefer to receive files in xml format via ftp, although metadata harvesting with OAI-PMH is not excluded. However, harvested material will most likely be in basic Dublin Core format, which is poor material for Exalead®'s refinement feature.
We defined a metadata scheme based on the DSpace scheme (qualified Dublin Core), but taking into account new needs (e.g. for Exalead®'s refinements) and ideas from similar repositories. Suggestions for additions are very welcome.

Updates to the OpenGrey database are done by batch upload of files, excluding direct deposit through the interface, contrary to institutional repositories[8].
Tools for participants will be made available through the website. It might be interesting to open a community space dedicated to partners. On a first meeting it was suggested to create an advisory board to discuss arising questions.

Outlook

Knowing than many documents referenced in the former SIGLE database have since been digitized or became available in electronic format, leads to another point for improvement. Our plans for future developments include the addition – if possible through batch upload – of links to the now available full text in the partner repository.
Another project planned with Greynet is to add persistent links to datasets in the bibliographic records, when they are available in a repository. The initiative might start with the next GL conference.

Glossary

Creative Commons: The Creative Commons copyright licenses and tools forge a balance inside the traditional "all rights reserved" setting that copyright law creates: http://creativecommons.org/licenses/
DSpace: DSpace is a software for academic, non-profit, and commercial organizations building open digital repositories: http://www.dspace.org/

Dublin Core Metadata: The Dublin Core Metadata Initiative (DCMI) is an open organization, incorporated in Singapore as a public, not-for-profit Company engaged in the development of interoperable metadata standards that support a broad range of purposes and business models: http://dublincore.org/specifications/

EAGLE: European Association for Grey Literature Exploitation**,** producer of the SIGLE database: http://en.wikipedia.org/wiki/European_Association_for_Grey_Literature_Exploitation
Exalead® is a global software provider in the enterprise and Web search markets: http://www.exalead.com/software/

Google Scholar: Google Scholar provides a simple way to broadly search for scholarly literature ...: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites: http://scholar.google.com/

Handle: The Handle System provides resolution services for unique and persistent identifiers of digital objects, and is a component of CNRI's Digital object (Corporation for National Research Initiatives® is a not-for-profit organization formed to undertake, foster, and promote research in the public interest) : http://www.handle.net/

OAI-PMH: The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH in the remainder of this document) provides an application-independent interoperability framework based on metadata harvesting*:* http://www.openarchives.org/OAI/openarchivesprotocol.htm

OpenDOAR: Directory of Open Access Repositories : The OpenDOAR service provides a quality-assured listing of open access repositories around the world: http://www.opendoar.org/find.php

OpenSIGLE: http://opensigle.inist.fr

SIGLE: System for Information on Grey Literature in Europe: SIGLE was an online, pan-European electronic bibliographic database and document delivery system for grey literature. http://en.wikipedia.org/wiki/SIGLE

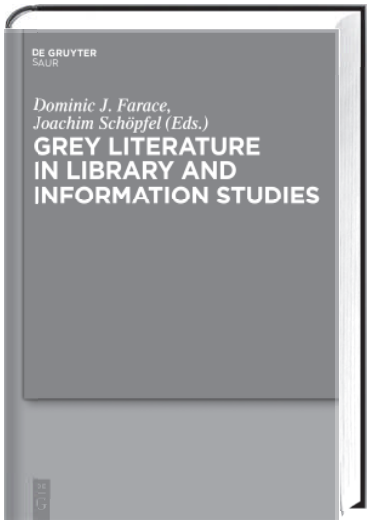
php/MyVisites: is a free and open source (GNU/GPL) software for websites statistics and audience measurements: http://www.phpmyvisites.us/

WorldWideScience.org: a global science gateway comprised of national and international scientific databases and portals: http://worldwidescience.org

References

[1] http://opensigle.inist.fr/handle/10068/697774
[2] the main effort relating to the identification and collection of documents
[3] mainly because it used standard DSpace features without further development
[4] older software based on simple log analysis would yield numbers at least four times higher
[5] http://opensigle.inist.fr/handle/10068/697753
[6] http://nrgl.techlib.cz/index.php/Main_Page
[7] E.g. NTK : http://opensigle.inist.fr/handle/10068/20
[8] However, a special development might allow partner administrators from partners to directly edit individual records.

https://doi.org/10.26069/greynet-2026-000.555-tgj



Dominic J. Farace and Joachim Schöpfel (Eds.)

GREY LITERATURE IN LIBRARY AND INFORMATION STUDIES

2010. vi, 282 pages
Hardcover RRP € 89.95 [D] / US$ 126.00. ISBN 978-3-598-11793-0
eBook RRP € 89.95 / US$ 126.00. ISBN 978-3-598-44149-3

The further rise of electronic publishing has come to change the scale and diversity of grey literature facing librarians and other information practitioners. This compiled work brings together research and authorship over the past decade dealing with both the supply and demand sides of grey literature. While this book is written with students and instructors of Colleges and Schools of Library and Information Science in mind, it likewise serves as a reader for information professionals working in any and all like knowledge-based communities.

CONTENTS

Introduction Grey Literature *(Farace and Schöpfel)*

Part I – Producing, Processing, and Distributing Grey Literature

Section One: Producing and Publishing Grey Literature
Chapter 1 Grey Publishing and the Information Market: A New Look at Value Chains and Business Models *(Roosendaal)*
Chapter 2 How to assure the Quality of Grey Literature: The Case of Evaluation Reports *(Weber)*
Chapter 3 Grey Literature produced and published by Universities: A Case for ETDs *(Južnič)*

Section Two: Collecting and Processing Grey Literature
Chapter 4 Collection building with special Regards to Report Literature *(Newbold and Grimshaw)*
Chapter 5 Institutional Grey Literature in the University Environment *(Siegel)*
Chapter 6 Copyright Concerns Confronting Grey Literature *(Lipinski)*

Section Three: Channels for Access and Distribution of Grey Literature
Chapter 7 Theses and Dissertations *(Stock and Paillassard)*
Chapter 8 Grey Documents in Open Archives *(Luzi)*
Chapter 9 OpenSIGLE – Crossroads for Libraries, Research and Educational Institutions in the Field of Grey Literature *(Farace, Frantzen, Stock, Henrot, and Schöpfel)*

Part II – Uses, Applications, and Trends in Grey Literature

Section Four: Applications and Uses of Grey Literature
Chapter 10 The driving and evolving Role of Grey Literature in High-Energy Physics *(Gentil-Beccot)*
Chapter 11 The Use and Influence of Information Produced as Grey Literature by International, Intergovernmental Marine Organizations: Overview of Current Research *(MacDonald, Wells, Cordes, Hutton, Cossarini, and Soomai)*
Chapter 12 Grey Literature in Karst Research: The Evolution of the Karst Information Portal, KIP *(Chavez)*
Chapter 13 Grey Literature Repositories: Tools for NGOs Involved in Public Health Activities in Developing Countries *(Crowe, Hodge, and Redmon)*

Section Five: Future Trends in Grey Literature
Chapter 14 Blog Posts and Tweets: The Next Frontier for Grey Literature *(Banks)*
Chapter 15 Assessing the Return on Investments in Grey Literature for Institutional Repositories *(Schöpfel and Boukacem)*
Chapter 16 e-Science, Cyberinfrastructure and CRIS *(Jeffery and Asserson)*
Chapter 17 Course and Learning Objective in the Teaching of Grey Literature: The Role of Library and Information Science Education *(Rabina)*

GL Transparency: Through a Glass*, Clearly† ‡

Keith G. Jeffery *(United Kingdom)* and Anne Asserson *(Norway)*

Abstract
GL (Grey literature, interpreted here as grey objects) is very heterogeneous in content, form and quality. Most GL objects evolve through a workflow. Some of these phases involve some form of evaluation or peer review, commonly internal within the management structure of an organisation and possibly involving external advice, including from 'friendly peers' via an e-preprint mechanism. Unlike white literature the evaluation process commonly is unrecorded and undocumented. This leads to accusations that grey literature lacks quality and transparency. This paper proposes how the GL community can overcome this – generally unfounded – accusation, building on our previous work.

A GL repository records the intellectual property of that organisation (2004). We have demonstrated that effective use of this resource requires that the metadata is formalised (1999, 2004) – more precisely in a CERIF-CRIS (Common European Research Information Format – Current Research Information System) (2005). The GL is then available in the context of the work of the organisation and/or its stakeholders managing strategy, evaluation, funding and cost-accounting, innovation and knowledge transfer and public information (2005). This provides user-evaluated assurance on the quality and relevance of the grey object. CERIF provides temporally-based relationships between grey objects (and white objects) thus recording evolution of the object during the workflow – hence provenance. This concept was further refined as 'Greyscape' (2007) and the technologies for interoperation – in order to provide the underpinning homogeneous access to the heterogeneous repositories - surveyed (2008). Efficiency of using CERIF was outlined in (2009). Using advanced hyperactive objects (2006) is postponed until the requirement is realised by the community.

CERIF-CRIS provides the capability for greater quality and transparency through novel methods of evaluating quality, provenance and review including Web2.0 recommender-type systems as well as conventional review mechanisms. CERIF-CRIS provides the way to overcome criticism of GL.

The key messages are:
1. *formal metadata associated with grey literature repositories improves relevance and quality;*
2. *transparency requires recording the workflow phases of a grey object within the context of a research information system;*
3. *a solution – CERIF – exists already which covers these requirements.*

Background

Previous Work

For more than two decades, the authors have worked on research information in the widest sense comprising information not only about grey literature (grey objects) but also all the outputs of research (products, patents, publications) and the context within which the research was done including projects, organizations, funding, persons, facilities, equipment, events. Within the GL community we have highlighted the issues as we see them:

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

* GLASS: Grey Literature Architecture for Sustainable Systems
† Clearly: "For now we see through a glass, darkly". The Bible: 1 Corinthians xiii, 12
‡ First published in the GL12 Conference Proceedings, February 2011

5. A 'from 10,000 metres altitude' view of the grey information landscape 'Greyscape' based on the hypothesis that grey literature is the foundation for the knowledge economy (Jeffery and Asserson 2007).
6. An analysis of interoperation architectures among research information systems 'INTEREST' (Jeffery and Asserson 2008).
7. A proposal that Grey Literature should be seen within the context of e-Science supported by a CERIF-CRIS (Jeffery and Asserson 2009).

The Requirement

Our work has convinced us of the need in the Grey literature community for two key technologies:

a) Metadata with formal syntax and declared semantics to allow reliable, scalable management and interoperation of grey resources;
b) Workflow within a research process context to minimize effort for the researcher, research manager, librarian or other knowledge worker, to record the provenance of a grey object and thus to increase accuracy, relevance and contextual awareness of the research information;

These technologies are required for many purposes in GL including – but not limited to – transparency. Transparency is defined in physics as the property of allowing light to pass through a material while more generally it implies openness, communication, and accountability. The latter meaning is used in this paper.

We contend that the currently widely-accepted metadata standards for GL – namely Dublin Core (DC) and (MARC) – are insufficient for the purposes of:

a) Discovery (relying on multilingual semantics over multicharactersets);
b) Management (utilising especially the formal syntax);
c) Utilisation (including security and privacy);
d) Understanding (relying on semantics);
e) Re-purposing (relying on both syntax and semantics);
f) Contextualising (utilising contextual metadata such as project, organisation);
g) Provenance (the stages through which the material has been);
h) Preservation/curation (for later re-use by future researchers);
i) Quality assessment (utilising the recorded workflow steps (provenance));

Without appropriate metadata transparency is lost (and the impact of the work recorded in the GL object is much reduced).

Further, we contend that unless GL material is collected in the context of a research workflow of services acting on the grey objects:

a) the threshold barrier to collection is high and discourages those producing the GL from providing the metadata (or even the source material);
b) associated contextual information is lost including any quality controls or peer review, or information allowing reputational judgement – thus transparency, so essential for confidence and trust in the information, is also lost;

We propose that both of these problems are overcome by use of a CERIF-CRIS.

The Hypothesis

The hypothesis is in three assertions:

1. formal metadata associated with grey literature repositories improves relevance and quality;
2. recording the workflow phases of a grey object within the context of a research information system provides provenance;
3. a solution – CERIF – exists already which covers these requirements.

Utilisation of this technology provides a GLASS (Grey Literature Architecture for Sustainable Systems) enabling GL users previously "seeing through a glass darkly" to see clearly.

Proposed Architecture

Introduction

The proposed GLASS – to achieve all the required aspects of a GL environment including transparency - consists of grey objects, metadata and services operating over a virtualised e-infrastructure based on GRIDs (as proposed in (Jeffery and Asserson 2009) or CLOUD technology (for a survey and analysis see (Schubert, Jeffery, Neidecker-Lutz 2010)) thus in

the same domain as that in which researchers do their other work. In this way activities associated with GL are not divorced from observation, experimentation, simulation or project management.

Grey Objects

It is expected that the grey objects will be heterogeneous (either in the local collection or the virtual collection obtained by accessing across heterogeneous distributed GL repositories) and of various (multi)media types. The only architectural problems concerning the objects are to ensure that appropriate services are available to utilise them (see list of functions in section 1.2). This implies rich metadata related to the objects to characterise the way in which they are utilised.

Metadata for Grey Objects

There are several classifications of metadata and multiple standards across many domains of scholarly research. The classification of metadata should relate to the purposes for which it will be utilised (through services available to the user) related to the object. For example schema metadata is used to assure integrity whereas descriptive metadata is used - among other purposes - for discovery.

Table 1: Metadata Kinds for Grey Objects Related to Services

SERVICE	METADATA	COMMENTS
Discovery	Descriptive	Multicharacterset, multilingual
Management	Schema Descriptive Restrictive Navigational Provenance Curation/Preservation	Depending on the management process different kinds of metadata are utilised
Utilisation	Schema Descriptive Restrictive Navigational	The schema metadata connects the object to the service assuring integrity, the descriptive metadata assures relevance and the restrictive metadata assures rights, security, privacy compliance
Understanding	Descriptive Provenance	The descriptive metadata assures relevance. The provenance metadata illuminates the evolution of the object
Re-Purposing	Schema Descriptive Restrictive Navigational Contextual Provenance Curation/Preservation	In order to re-use a grey object as much information about it as possible is required to assure that the re-use is valid.
Contextualising	Contextual	For example placing the object in the context of a research project, or related to a research facility
Provenance	Provenance	One aspect of quality
Preserving / Curating	Schema Descriptive Restrictive Navigational Contextual Provenance Curation/Preservation	All metadata is needed to allow re-purposing at a later time when the grey object creator may be unavailable
Quality Assessment	Schema Contextual Provenance	The schema metadata provides integrity, provenance metadata describes the object evolution and contextual metadata covers the research context

Clearly there are advantages if the metadata for grey objects is stored within one standard structural environment. CERIF provides such an environment covering all the kinds of metadata outlined above, except schema which – by definition – relates to the conceptual, logical and physical representation of the object within the hosting environment. The availability of this rich metadata for grey data objects assures transparency.

Services

Services are executed to fulfil the requirements of the end-user. Services, themselves, require metadata in the same way as grey objects. Services need to be discovered, managed etc. Services can be :

a) object-independent i.e. generic processes that act on any data or
b) object-dependent i.e. including and enclosing the object(s) together with the processes.

Examples of (a) are the relational algebra operators (select, project, difference, union, join) which then act on any relational table(s) whatever the data stored in those tables. Examples of (b) include currency conversion services where the current exchange rate table is incorporated within the service and its associated atomic processes. There is an argument that composed services should themselves be described by metadata and contain both processes and objects, each of which is also described by metadata.

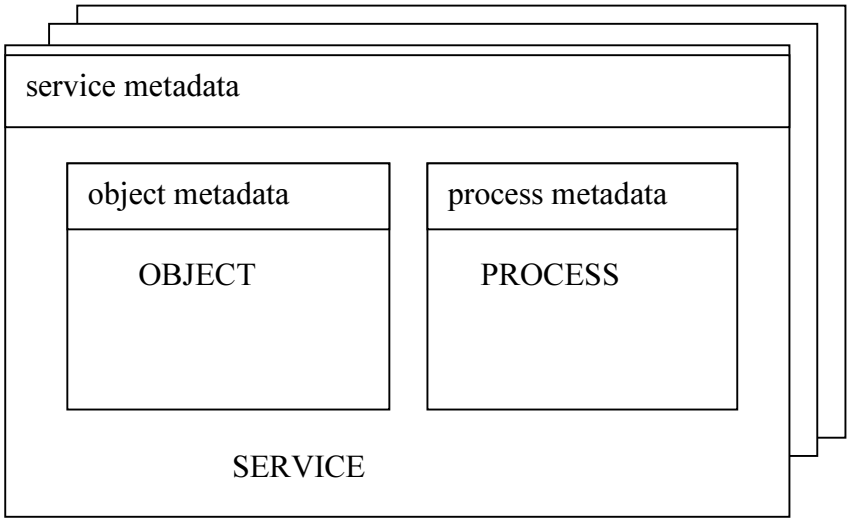


Figure 1: Service, Object, Process Metadata

However, whereas objects can be collected together in collections (usually as a set of similar objects based on some parameters e.g. all grey objects generated in 2010 relating to global warming and climate change) services also may be composed – that is groups of atomic services are linked together in some ordered fashion to execute the requirements of the end-user. A simple sequence of services (e.g. select, count) is a simple workflow. A more complex workflow has decision points and branches. However, for reasons of performance and resilience – especially in a heterogeneous distributed environment – the composition can include multiple parallel tracks of workflow with replicated services. It is necessary that each service can be executed anywhere – depending on requirements of performance, security etc – which demands that the services are mobile – that is the program code can be moved to the locus of execution. This leads to the requirement for self-organising (composing, managing, optimizing) services because the complexity of matching the requirement (including non-functional aspects such as performance, security, adherence to a service level agreement or quality of service) to the execution environment (distributed, heterogenous, parallel, multi-tenanted) is too great and too dynamic for human management.

Within the architecture task group of euroCRIS, a set of services for any CERIF-CRIS is being discussed.

Metadata for Services

The services provided require metadata in order for them to be utilised and – more importantly – for them to be utilised correctly.

1. Schema metadata: analogous to the schema controlling integrity in a data object is used to assure integrity in the service particularly in the parameters and input/output declarations;
2. Navigational metadata: analogous to that for data objects is used to locate the service:

3. Descriptive metadata: analogous to that for data objects is used to discover the service and then (together with the schema and restrictive metadata) to assure fitness for purpose;
4. Restrictive metadata: analogous to that for data objects assures enforcement fo non-functional properties of the service such as performance, security, privacy, rights management, (micro-)payment for usage;
5. Provenance metadata: analogous to that for data objects records the transition states of the service as it evolves;
6. Curation/Preservation metadata: analogous to that for data objects records the additional information require to assure (a) preservation of the service code and documentation (specification) over time e.g. through media conversion and evolution; (b) curation such that the purpose and characteristics may be understood in future time;
7. Contextual metadata: analogous to that for data objects describing how the service fits within a context of other research information such as projects, organizations, facilities, equipment etc. CERIF is the preferred standard for use here.
   The availability of this rich metadata for services in the GL domain assures transparency.

Conclusion

The proposed GLASS architecture achieves transparency through several mechanisms:

1. encouraging the provision of full metadata using CERIF to cover all aspects of the grey data object thus maximizing the potential utilisation and providing information relating to integrity and quality;
2. encouraging the provision of full metadata using CERIF to cover all aspects of services thus maximizing the potential utilization (including in composed services) and providing information relating to integrity and quality;
3. through CERIF defining metadata with formal syntax (for reliable computer processing) and declared semantics (for computer or human understanding);
4. through CERIF providing a data model which records the date/time interval associated with any relationship between two base entities. This provides automatically a provenance track and also can be used for non-functional aspects such as security, privacy, rights restrictions;

References

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

(Asserson and Jeffery 2005) Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' The Grey Journal volume 1 number 1: Spring 2005 TextRelease/Greynet ISSN 1574-1796 pp5-8

(CERIF) www.eurocris.org/cerif

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

(DSpace) DSpace repository product homepage http://www.dspace.org/

(ePubs) ePubs repository product homepage http://epubs.cclrc.ac.uk/

(eprints) ePrints repository product homepage http://www.eprints.org/

(Fedora) Fedora repository product homepage http://www.fedora.info/

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

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

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

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

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

(Jeffery 2005) K G Jeffery CRISs, Architectures and CERIF CCLRC-RAL Technical Report RAL-TR-2005-003 (2005)

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

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

(Jeffery and Asserson 2006a) Keith G Jeffery, Anne Asserson: 'CRIS Central Relating Information System' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond he Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp109-120 Leuven University Press ISBN 978 90 5867 536 1

(Jeffery and Asserson 2006b) Keith G Jeffery, Anne Asserson: 'Supporting the Research Process with a CRIS' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond the Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp 121-130 Leuven University Press ISBN 978 90 5867 536 1

(Jeffery and Asserson 2006c) Keith G Jeffery, Anne Asserson: 'Hyperactive Grey Objects' Proceedings Grey Literature 8 Conference (GL8), New Orleans, December 2006; TextRelease; ISBN 90-77484-08-6. ISSN 1386-2316 ; No. 8-06-X

(Jeffery and Asserson 2007) Keith G Jeffery, Anne Asserson: 'Greyscape' Opening Paper in Proceedings Grey Literature 9 Conference Antwerp (GL9) 10-11 December 2007 pp9-14; Textrelease, Amsterdam; ISSN 1386-2316

(Jeffery and Asserson 2008) Keith G Jeffery, Anne Asserson: 'INTEREST' Proceedings Grey Literature Conference Amsterdam 8-9 December 2008 in Tenth International Conference on Grey Literature : Designing the Grey Grid for Information Society, 8-9 December 2008, Science Park Amsterdam, The Netherlands ed. by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2009. GL-Conference series, ISSN 1386-2316; No. 10. - ISBN 978-90-77484-11-1.

(MARC) http:// http://www.loc.gov/marc/

(PURE) http:// http://www.atira.dk/en/pure/

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

Schubert, Jeffery, Neidecker-Lutz 2010 'The Future of Cloud Computing' http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf .

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

(Z39.50) http:// http://www.loc.gov/z3950/agency/

https://doi.org/10.26069/greynet-2026-000.555-tgj

Invenio: A Modern Digital Library System for Grey Literature*

Jérôme Caffaro and Samuele Kaplun (Switzerland)

Abstract
Grey literature has historically played a key role for researchers in the field of High-Energy Physics (HEP). Consequently CERN (European Organization for Nuclear Research) as the world's largest particle physics laboratory has always been facing the challenge of distributing and archiving grey material. Invenio, an open-source repository software, has been developed as part of CERN's institutional repository strategy to answer these needs.
In this document we describe how the particular context of grey literature within the HEP community shaped the development of Invenio. We focus on the strategies that have been established in order to process grey material within the software and we analyse how it is used in a real production environment, the CERN Document Server (CDS).

Introduction

Background

CERN The European Organization for Nuclear Research in Geneva is the world's largest particle physics laboratory. Originally founded in 1954 by 12 European countries, CERN has established a solid reputation in scientific research throughout history. CERN is currently run by 20 European member states with over 40 additional participating observers (states and organizations). About 8,000 scientists from around 580 universities come to CERN to work on their research. The main current research program at CERN is the LHC (Large Hadron Collider), the largest (a 27km ring of superconducting magnets) and most powerful accelerator, smashing particles together to understand the basic constituent of matter. Over a thousand publications are published yearly by CERN scientists in established journals.

HEP Community The High Energy Physics (HEP) community is estimated to have about 20,000 scientists. It essentially comprises researchers working in the major particle physics laboratories around the world such as CERN, Desy (Germany) Fermilab (USA), SLAC (USA) and KEK (Japan).

Invenio

Invenio is an integrated digital library system [1] originally developed at CERN to run the CERN Document Server (CDS). It is currently one of the largest institutional repositories worldwide. It was started over 15 years ago and has matured through many release cycles. Invenio is a GPL2 Open Source project based on an Apache/WSGI+Python+MySQL architecture. Its modular design enables it to serve a wide variety of requirements, from a multimedia digital object repository, to a web journal, to a fully functional digital library. The development strategy used to implement Invenio ensures that it is flexible in every layer. Being based on open standards such as MARCXML and OAI-PMH 2.0 its interoperability with other digital libraries is guaranteed. Having been originally designed to cope with the CERN requirements for digital object management, Invenio is suitable for middle-to-large scale digital repositories (100K~10M records).

Grey Material at CERN

Invenio software was born in a rich grey literature producing environment [2]. One early major impact on the HEP community, and by consequence on the development of Invenio was the definition of a strong policy towards the dissemination of the work done at CERN. Indeed the convention that established CERN in 1954 states that "*[...] the results of its experimental and theoretical work shall be published or otherwise made generally available*" [3]. The implied openness of this mission forged a strong idea of responsibility for the community to give access not only to published documents, but also to additional material produced by the organization. CERN being an international organization, involving collaborations with a large number of universities and institutions, efficient sharing of information was a primary concern not only for scientific results, but also for all the material necessary for good coordination and proper running of the experiments:

* First published in the GL12 Conference Proceedings, February 2011

engineering drawings, technical reports, notes, etc. containing important scientific or technical data, but not suitable for publication in journals.

An important trend that took off among HEP researchers more than 50 years ago was the habit of mailing to their peers printed copies of their work at the time of submission to journals [4], reducing by several months the access to results of a possibly major importance in the context of the laboratory: machines and tools built for the CERN experiments being giant and technically advanced prototypes, they require several iterations of optimizations which can be performed through the analysis of early experimental results. Reducing the duration of these cycles was necessary to help keeping the cost of the experiments as low as possible.
Early experimental results would also be used as input by theoretical physicists in order to refine existing theories or build new ones, and suggest new areas of study for experimental physics [5].

Another early concern expressed by the researchers at CERN was simply related to the physical access to grey literature. It is common for several thousands of scientists and engineers to work on the same experiment due to its complexity and size. A geographical distribution of the experts is inevitable, considering both the country of origin of these experts and location on the experiment(s) site(s) for practical reasons. Coordination of such large projects is only possible through formal, written forms. That particular need of giving access to a large amount of information in an electronic way resulted in the creation of the World Wide Web in 1989, as a proposal by Tim Berners Lee [6].

In the last two decades, HEP continued to pioneer solutions in scholarly communications. SPIRES, the SLAC (Stanford Linear Accelerator) literature database became in 1991 the first web server in the USA and the first online database in the world [7]. The same year arXiv.org (named at the time "LANL preprint archive") opened its web frontend, first as repository of physics preprints before expanding to other fields of science. In 1993 CERN released its preprints database on the web as an early version of the Invenio software, with the initial goal of fulfilling CERN's needs of access to grey literature.

Invenio for Grey Literature

In this section we review the strategies adopted in Invenio to support the management of grey literature. Examples of applications to real production systems running Invenio are given. In particular numbers given in the following paragraphs are updated statistics of the CERN Document Server (powered by Invenio) for November 2010.

MARCXML as core metadata format

MARCXML is the core bibliographic metadata format of Invenio. It offers all the required flexibility to model a great variety of digital assets. Being a library standard, MARCXML offers the advantage of being well-known by professional librarians, giving a unique chance for grey material to be curated by the institutional library team. The very same tools used to manage published material can be used to process grey material, ensuring higher quality data and helping grey literature find its way more to standard institutional processes more easily.

Flexible metadata-formatting layer

A flexible metadata formatting layer in combination with the MARCXML format allows the visualization of practically any digital asset within Invenio. An accessible HTML-like markup offers the opportunity for librarians to define the display of the managed records. Additional support for XSLT at the level of the formatting layer enables easy conversion from MARCXML to other XML flavor formats in a standard way. CDS uses 122 different formatting templates, approximately half of them being used to prepare search results output, and the other half providing detailed information about the records. The templates deal with standard preprint objects, as well as video, audio or photo content. Combined with a customizable collection tree, it is possible to offer subject-based "portals" regardless of the actual type of contained material.

Customizable workflow engine

The submission system of Invenio lets administrators configure their own customized workflows. The framework offers the tools to create web front-ends for users to submit data (metadata and files), and an extensible set of functions to process the collected data. Typical workflows result in the creation of a new record in the repository. Thanks to the mapping of the collected data to MARCXML, the flexibility offered by the submission system of Invenio regarding the type of supported data can be extended to the archival of this data. Submissions of Invenio can also be the starting point for subject specific jobs such as

OCR (Optical Character Recognition) for scanned documents, and image downsizing for the creation of web versions. 91 different workflows are currently maintained on the CERN Document Server, some very similar to the basic workflow described above and other implementing complex reviewing and approval systems implicating thousands of different users. An average of more than 30 documents is submitted per day (less during week-ends) through these web-based workflows (300 documents per day when considering alternative input methods).

Collaborative tools
Invenio supports the creation of user groups (local or derived from the institution identity management system) and "baskets" letting users share information in a controlled, targeted manner. This feature is particularly useful in order to provide a community-based selection of unpublished documents at a quality that a ranking algorithm cannot match. Shared baskets can then offer a fast changing community-based hierarchical structure of data that Invenio main navigable collection tree cannot provide. In November 2010, CDS counted 6,245 non-empty baskets set up by 4,575 distinct users, covering about 10% of the whole archive. 6% of the baskets were shared among several users. Invenio also features basic commenting and reviewing capabilities enabling a better understanding of the quality of the material. Consequently it also archives information about the documents which in the past was wrongly regarded as transient information only. The most active collection on CDS, regarding the number of comments, gets an average of about 30 comments per day through its built-in commenting system, usually on recent work being under peer-review at CERN.
Access control also plays an important role in helping the ingestion of grey material: providing an adequate, secure and restricted collaborative workspace for draft documents is an incentive for users to move their workflow to a central server, hence giving more opportunity for the final document to be made publicly (or not) available. Indeed Invenio can accompany documents through their life cycle, thanks to the integration of an advanced role-based access control system into the flexible workflow engine.
To bring a better awareness of the quality of the material to users and help with the discovery of documents of possible interest, Invenio display recommendation based on document usage statistics (*"People who viewed this page also viewed..."*) and features download/citation history graphs.

Search engine A major concern for large repositories of grey material is to provide an efficient way to retrieve new and archived material. Invenio includes a very fast search engine optimized for large repositories (millions of documents) on simple infrastructures, combining metadata and fulltext search in a simple Google-like query language. Advanced users are also given the opportunity to perform advanced queries, such as *find document written by Ellis from years 2000 to 2010, mentioning "higgs boson" in the fulltext, referring to documents written by Randall, and cited more than 50 times*. The CERN Document Server serves about 25,000 queries per day, for an archive of about 1 million records. Retrieved documents can be ranked according to several techniques, such as "word similarity" ranking or "citation-graph" [8] based ranking etc. in order to accommodate to the type of searched material: for example a researcher new to some subject is more likely to search for general reference documents while an expert might rather be looking for all new material in his field of interest.

The Invenio search engine technology is at the core of many functionalities offered by the software. For example combined with the flexible metadata-formatting layer, it can provide personalized search-based RSS feeds or email alerts: new results to some specific search queries such as the sample one mentioned above can be sent periodically to the subscribers. This has proven to be an essential functionality for CERN physicists in need of the latest information on some very specific topics. CDS counts more than 12,000 RSS subscriptions set up by 3000 distinct users (IP-based). 2417 emails alerts have also been set up by 1615 users.

Other use cases of the search engine include the suggestion of documents similar to a given one, or the creation of the bibliography (BibTEX) of a given author, personalized podcasts, etc.

Interoperability Invenio implements standard protocols to help the ingestion and dissemination of documents. Though these protocols are usually independent of document-type, they can still suffer from the conversion process usually occurring to ensure that repositories use a common language. For example OAI-PMH is able to support any document type, but most repositories only support the Dublin Core schema, hence

narrowing down the possibility to use this protocol in some scenarios. Invenio is able to export and import any metadata format in OAI-PMH thanks to the underlying layers supporting custom conversion templates. For example OAI-PMH is used at CERN to feed an installation of Invenio with conference and meeting objects coming from the institutional conference management system Indico. Another example of usage of OAI-PMH in Invenio is in the context of the OPENAIRE project, which is planning to exchange usage statistics among participating repositories through this protocol.

Integrated digital library Invenio is a multi-purpose repository software: not exclusively designed for grey material, it offers the advantage of being a solution for the common needs of a library. It results in lower infrastructure maintenance costs by grouping several library services and processes on a single server. Reusing the same technology and concepts for these different services is also reducing the learning curve to master the necessary tools. It becomes much easier to justify the cost of supporting grey material within the institution.

Conclusions

Invenio is a well-established open source repository software. The context in which the software was conceived and then further developed has played an important role in defining a core set of features suitable for the ingestion, processing and distribution of grey material.

The performance and flexibility of the software has led to its adoption in a variety of scenarios, strengthening the will to drive the development efforts towards an increased support for grey material.

References

[1] Invenio official website.
http://invenio-software.org/
(Last visited on 19 November 2010)

[2] L. Goldschmidt-Clermont, Communication Patterns in High-Energy Physics; High Energy Physics Libraries Webzine, issue 6, March 2002.
http://library.web.cern.ch/library/Webzine/6/papers/1/
(Last visited on 18 November 2010)

[3] Convention for the Establishment of a European Organization for Nuclear Research, Paris, 1st July, 1953
http://council.web.cern.ch/council/en/Governance/Convention.html
(Last visited on 18 November 2010)

[4] Anne-Gentil Beccot et al., Information Resources in High-Energy Physics: Surveying the Present Landscape and Charting the Future Course, J.Am.Soc.Inf.Sci.60, 150-160, (2009)

[5] Anne-Gentil Beccot, How do High Energy Physics scholars search their information?, Grey J. 4, 1 (2008).

[6] Tim Berners-Lee Information Management: A Proposal CERN-DD-89-001-OC, 1989
http://cdsweb.cern.ch/record/369245

[7] L. Addis, Brief and Biased History of Preprint and Database Activities at the SLAC Library, 1962-1994.
http://www.slac.stanford.edu/spires/papers/history.html
(Last visited on 19 November 2010)

[8] Ludmila Marian et al., Citation graph based ranking in Invenio. LNCS (Research and Advanced Technology for Digital Libraries) 6273 (2010)

New Password Protected Access to LIS Serials

TextRelease now offers web-based access to serial publications on grey literature. Libraries and research centers with serial holdings in the field of library and information science, can now acquire password protected access to both current as well as back volumes of

The Grey Journal, International Journal on Grey Literature, ISSN 1574-180X

TGJ Volume 7, Number 1, Spring 2011	Transparency in Grey Literature
TGJ Volume 7, Number 2, Summer 2011	System Approaches to Grey Literature
TGJ Volume 7, Number 3, Autumn 2011	*(To be announced)*
TGJ Volume 6, Number 1, Spring 2010	Government Alliance to Grey Literature
TGJ Volume 6, Number 2, Summer 2010	Shared Strategies for Grey Literature
TGJ Volume 6, Number 3, Autumn 2010	Research on Grey Literature in Europe
TGJ Volume 5, Number 1, Spring 2009	Paperless Initiatives for Grey Literature
TGJ Volume 5, Number 2, Summer 2009	Archaeology and Grey Literature
TGJ Volume 5, Number 3, Autumn 2009	Trusted Grey Sources and Resources
TGJ Volume 4, Number 1, Spring 2008	Praxis and Theory in Grey Literature
TGJ Volume 4, Number 2, Summer 2008	Access to Grey in a Web Environment
TGJ Volume 4, Number 3, Autumn 2008	Making Grey more Visible
TGJ Volume 3, Number 1, Spring 2007	Grey Standards in Transition and Use
TGJ Volume 3, Number 2, Summer 2007	Academic and Scholarly Grey
TGJ Volume 3, Number 3, Autumn 2007	Mapping Grey Resources
TGJ Volume 2, Number 1, Spring 2006	Grey Matters for OAI
TGJ Volume 2, Number 2, Summer 2006	Collections on a Grey Scale
TGJ Volume 2, Number 3, Autumn 2006	Using Grey to Sustain Innovation
TGJ Volume 1, Number 1, Spring 2005	Publish Grey or Perish
TGJ Volume 1, Number 2, Summer 2005	Repositories - Home2Grey
TGJ Volume 1, Number 3, Autumn 2005	Grey Areas in Education

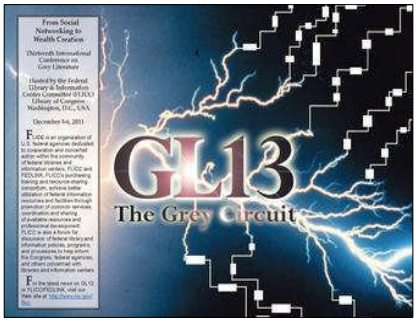
International Conference Proceedings on Grey Literature, ISSN 2211-7199

GL12	Prague, Czech Republic, 2010	Grey Tech Approaches to High Tech Issues
GL11	Washington D.C., United States, 2009	The Grey Mosaic, Piecing It All Together
GL10	Amsterdam, Netherlands, 2008	Designing the Grey Grid for Information Society
GL9	Antwerp, Belgium, 2007	Grey Foundations in Information Landscape
GL8	New Orleans, United States, 2006	Harnessing the Power of Grey
GL7	Nancy, France, 2005	Open Access to Grey Resources
GL6	New York, United States, 2004	Work on Grey in Progress
GL5	Amsterdam, Netherlands, 2003	Grey Matters in the World of Networked Information

Note: Current subscribers as well as GreyNet members are eligible for a 20% reduction on all back volumes. Ask for a free price quote today, **info@textrelease.com**

https://doi.org/10.26069/greynet-2026-000.555-tgj

Thirteenth International Conference on Grey Literature



Library of Congress, Washington D.C. USA, 5-6 December 2011

The Grey Circuit
From Social Networking to Wealth Creation

Conference Program

DAY ONE

9:00-10:30

Opening Session

Chair, *Blane K. Dessy, Director FLICC-FEDLINK; Library of Congress, United States*

Welcome Address *(Including)* **A Discussion on Legal Aspects of Grey Literature** *Blane Dessy, Matthew Braun, Law Library of Congress, United States and Joachim Schöpfel, University of Lille, France*

Keynote Address ***Jens Vigen,*** *Director CERN Library, European Organization for Nuclear Research, Switzerland*

Opening Paper **Science-Forums.net: A Platform for Scientific Sharing and Collaboration** *Marilyn J. Davis, OSTI-DOE and Lance Vowell, IIA Inc., United States*

11:00-12:30

Session One – Social Networking

Chair, *Joachim Schöpfel, University of Lille, France*

Social Networking: Product or Process and What Shade of Grey?
Julia Gelfand and Anthony Lin, UCI Libraries, United States

Knowledge Communities in Grey *Claudia Marzi, Institute of Computational Linguistics, CNR, Italy*

Using Social Media to Create Virtual Interest Groups in Hospital Libraries *Yongtao Lin, Tom Baker Knowledge Centre and Kathryn M.E. Ranjit, Health Information Network Calgary; Peter Lougheed Knowledge Centre, Canada*

NYAM's Grey Literature Report: How Social Networking Adds Value for Users and the Grey Literature Community *Janie Kaplan, Lea Myohanen, E. Taylor, Ying Jia, L. Keith, and Lisa Genoese, NYAM, United States*

1:30-2:30

Introductions to Posters

Chair, Robin Harvey, *FLICC-FEDLINK; Library of Congress, United States*

[1]**Library of Congress Special Collection … African American Experience** *Otis D. Alexander, United States* [2]**Dependency on Regional Libraries for Grey Literature** *N. Chowdappa, L. Usha Devi, and C.P. Ramasesh, India* [3]**Central Registry … Slovak Universities** *Marta Dušková and Ľudmila Hrčková, Slovak Republic* [4]**Information-Seeking Behaviour of Slovenian Researchers** *Primož Južnič and Polona Vilar, Slovenia* [5]**Grey Literature at the Center for the History of Psychology** *Jodi Kearns, Cathy Faye, and Lynn Willis, United States* [6]**GreyNet International, A Back Office Report** *Dominic Farace, Netherlands* [7]**FLICC-FEDLINK, Federal Library Information Network** *Robin Harvey, United States* [8]***FedGrey, A Working Group on Grey Literature*** *Blane K. Dessy, United States,* [9]**Grey literature between tradition and innovation …** *Gabriella Pardelli, Sara Goggi, and Manuela Sassi, Italy* ▶

2:30-4:30

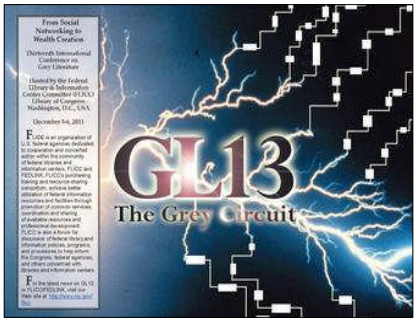
Poster Session and Sponsor Showcase

▶ *Twenty-five posters will be accommodated in this session. If you're keen to present, please contact TextRelease.*

CONFERENCE RECEPTION **5:00-6:30**

https://doi.org/10.26069/greynet-2026-000.555-tgj

Thirteenth International Conference on Grey Literature



Library of Congress, Washington D.C. USA, 5-6 December 2011

The Grey Circuit

From Social Networking to Wealth Creation

Conference Program

DAY TWO

9:00-11:00

SESSION TWO – SPECIAL COLLECTIONS

Chair, Stefania Biagioni, *Institute of Information Science and Technologies; National Research Council, Italy*

GeoStoryteller: Taking grey literature to the streets of New York
Debbie Rabina and Anthony Cocciolo, Pratt Institute, School of Information and Library Science, United States

Acquisition and distribution of technical reports and conference proceedings on science and technology in Korea *Seon-Hee Lee, Soon-Young Kim, Beom-Jong You, and Heeyoon Choi, KISTI, South Korea*

Describing Geography manifested through Grey Literature
Pete Reehling, University of South Florida, United States

Site Visit of the Geography and Map Division *John R. Hébert, Library of Congress, United States*

11:30-1:00

SESSION THREE – OPEN ACCESS AND WEALTH CREATION

Chair, *Seon-Hee Lee, Korea Institute of Science and Technology Information, South Korea*

Open Is Not Enough: A case study on grey literature in an OAI environment *Joachim Schöpfel, University of Lille 3, Isabelle Le Bescond, University of Lille 1, and Hélène Prost, INIST-CNRS, France*

Grey literature matters: The role of grey literature as a public communication tool in risk management practices of nuclear power plants *Cees de Blaaij, Library of Zeeland, Netherlands*

Management of Obsolete GL in Engineering Research Institutions *C.P. Ramasesh, University of Mysore, N. Chowdappa, BMS College of Engineering, L. Usha Devi, Bangalore University and V.R. Shyamala, P.U. College, India*

Audit DRAMBORA for trustworthy repositories: A Study Dealing with the Digital Repository of *GL*
Petra Pejšová, National Technical Library, Czech Republic; Marcus Vaska, Health Information Network Calgary, Canada

2:00-3:30

SESSION FOUR – DATA FRONTIERS

Chair, *Janie Kaplan, New York Academy of Medicine, United States*

Federal Information System on GL in Russia: A new stage of development in digital and network environment *Aleksandr V. Starovoitov, Aleksandr M. Bastrykin, Anton I. Borzykh, and Leonid P. Pavlov, CITIS, Russia*

Enhancing diffusion of scientific contents: Open data in Repositories *Daniela Luzi, Rosa Di Cesare, and Marta Ricci, IRPPS-National Research Council; Roberta Ruggieri, Senato della Repubblica, Italy*

Research product repositories: Strategies for data and metadata quality control *Luisa De Biagi, Roberto Puccinelli, Massimiliano Saccone, and Luciana Trufelli, National Research Council, Italy*

Linking full-text GL to underlying research and post-publication data: An Enhanced Publications Project
Dominic Farace and Jerry Frantzen, GreyNet, Netherlands; Christiane Stock, INIST-CNRS, France; Laurents Sesink, Data Archiving and Networked Services, Netherlands and Debbie Rabina, Pratt Institute, United States

3:30-4:00

CLOSING SESSION – REPORTS FROM CHAIRPERSONS, CONFERENCE HANDOFF, AND FAREWELL

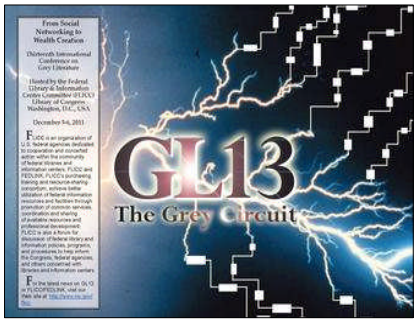
Chair, Blane K. Dessy, FLICC-FEDLINK, United States and Dominic J. Farace, GreyNet, Netherlands

POST-CONFERENCE TOUR OF THE JEFFERSON BUILDING **4:00-5:00**

GL13 Program and Conference Bureau
TextRelease
Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com · **conference@textrelease.com**
Tel/Fax +31-20-331.2420

https://doi.org/10.26069/greynet-2026-000.555-tgj

Thirteenth International Conference on Grey Literature



Library of Congress, Washington D.C. USA, 5-6 December 2011

The Grey Circuit
From Social Networking to Wealth Creation

Registration Form

The Conference Registration Fee includes attendance during the Sessions, a copy of the GL13 Program and Abstract Book, the full-text of the Conference Papers and PowerPoint Slides, as well as the conference pouch and participant badge. Also included are lunches, refreshments during the morning and afternoon breaks, the Inaugural Reception, and post-conference tour.

Payment after October 1, 2011 add a US$ 75 surcharge

☐ **US$ 595** Participant Fee ☐ **US$ 495** Author Fee ☐ **US$ 350** Day Pass ☐ **US$ 175** Student Fee

I work for a U.S. Government Agency and am entitled to 20% reduction on the fee: ☐

I am a *GreyNet* Member and am entitled to 20% reduction on the conference fee: ☐

Registrant's Name: _____

Organization: _____

Address/P.O. Box: _____

Postal Code/City/Country: _____

Tel/Fax/Email:

Please Check One of the Boxes below for the Method of Payment:

❑ Direct bank transfer to TextRelease, Account 3135.85.342 - Rabobank, Amsterdam, Netherlands
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to GL13 and Registrant's Name

❑ MasterCard ❑ Visa Card ❑ American Express

Card No. _____ Expiration Date: _____

If the name on the credit card is not that of the registrant, print the name that appears on the 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.***

GL13 Program and Conference Bureau
TextRelease

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

https://doi.org/10.26069/greynet-2026-000.555-tgj

Asserson, Anne **99**
Anne Asserson is presently representing UiB in the national group that is implementing a new national research documentation system, FRIDA. She has also participated in The CORDIS funded European-wide project on " Best Practice" 1996 . She was a member of the working group set up 1997 that produced the report CERIF2000 Guidelines (1999) www.cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model www.eurocris.org. Anne Asserson is a member of the Best Practice Task Group. anne.asserson@fa.uib.no

Di Cesare, Rosa **81**
Rosa Di Cesare is responsible for the library at the Institute for research on populations and social policies of the National Research Council (CNR). She worked previously at the Central library of CNR where she became involved in research activities in the field of Grey literature (GL) as member of the Technical Committee for the SIGLE database. Her studies have focused on the use of GL in scientific publications and recently on the emerging models of scholarly communication (OA and IR). r.dicesare@irpps.cnr.it

Henrot, Nathalie 93
Nathalie Henrot graduated in History, then in Information Sciences from the University of Tours in 1988. She has been working for the INIST-CNRS for seventeen years, more specifically at the Monographs & Grey Literature Section from 1993, for congress proceedings acquisition. She is now the user administrator in the OpenSIGLE project. Email: henrotn@inist.fr

Jeffery, Keith G. **99**
Keith Jeffery is currently Director, IT and International Strategy of STFC (Science and Technology Facilities Council), based at Rutherford Appleton Laboratory in UK. Keith is a Fellow of both the Geological Society of London and the British Computer Society. He is a Chartered Engineer. He is an Honorary Fellow of the Irish Computer Society. He is president of euroCRIS (www.eurocris.org) and of ERCIM (www.ercim.org) and holds three honorary professorships. He has extensive publications and has served on numerous programme committees and research grant review panels. He has particular interests in 'the research process' and the relationship of hypotheses, experiments, primary data and publications based on research in information systems, knowledge-based systems and metadata. Email: k.g.jeffery@rl.ac.uk

Luzi, Daniela **81**
Daniela Luzi is researcher of the National Research Council at the Institute of research on populations and social politics. Her interest in Grey Literature started at the Italian national reference centre for SIGLE at the beginning of her career and continued carrying out research on GL databases, electronic information and open archives. She has always attended the International GL conferences and in 2000 she obtained an award for outstanding achievement in the field of grey literature by the Literati Club. Email: d.luzi@irpps.cnr.it

Mynarz, Jindřich **65**
Jindřich Mynarz has got a bachelor's degree in Library and information science at the Institute of Information Studies and Librarianship, Charles University in Prague, and he continues with New media studies accredited at the same university. He works at the Development of electronic services department at the National Technical Library in Prague, Czech Republic. The main focus of his work is on library data and their transformation to more web-compatible data models and their exposing in conformance with linked data principles. Other areas of interest include knowledge organisation systems, particularly in the ways in which they can be used in linked data setting, and in user interfaces that can be built to access library data. Email: jindrich.mynarz@techlib.cz

Pejšová, Petra **72**
Petra Pejšová studied information science and librarianship at Charles University. She works as an information specialist in the State technical Library, Czech Republic. Actually she is leading a project Digital Library for Grey Literature – Functional model and pilot. Email: petra.pejsova@techlib.cz

Ricci, Marta **81**
Marta Ricci has an undergraduate degree in Humanities and a Master degree in Library Science from the University of Rome "Tor Vergata" with a thesis on bibliometric tools and citation analysis. She had an internship experience in the library of the Italian Prime Minister's office (Chigi's Library), where she was responsible for the Inventory of part of the library collections. Currently she is collaborating with the library of the Institute for Research on Population and Social Policies of the Italian National Research Council (CNR), in the field of Grey Literature. Email:biblio.irpps@irpps.cnr.it

Ruggieri, Roberta **81**
Roberta Ruggieri is librarian at the Senate of the Republic where she is responsible for the supervision of a digitalization project on Senate parliamentary print documents for the I to X Legislature. Her activity in managing digitalization project also includes document addition and classification in the electronic Senate catalogue. From 2004 she has been collaborating with the Institute for research on populations and social policies of the National Research Council (CNR) in research activities related to the field of Grey literature and Institutional repositories. Email: biblio.irpps@irpps.cnr.it

Škuta, Ctibor **65**
Ctibor Škuta is working in the Department of Polythematic Structured Subject Heading System (PSH) at the National Technical Library, Czech Republic. His tasks mainly involve the automation of processes related to the administration of PSH and cooperation on other projects with the Development of Electronic Services Department. At the same time, Škuta is studying Applied Informatics in Chemistry at The Institute of Chemical Technology in Prague. His professional interests are programming (Python, Java, XML technologies), data mining, and semantic web. Email: ctibor.skuta@techlib.cz

Stock, Christiane **93**
Christiane Stock is the Head of the Monographs and Grey Literature service at INIST, in charge of the repositories LARA (reports), mémSIC (master's theses in information sciences) and OpenSIGLE. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). She is member of the AFNOR expert group who prepared the recommended metadata scheme for French electronic theses (TEF). Email: christiane.stock@inist.fr

Vaska, Marcus **72**
Marcus Vaska is a librarian for the Physician Learning Program (PLP), a new collaborative initiative between the Universities of Alberta and Calgary, funded via an Alberta Medical Association (AMA) trilateral agreement. Marcus is responsible for assisting physicians in their research, and addressing their perceived and unperceived learning needs. Prior to this position, he was a librarian at the University of Calgary's Health Sciences Library. Marcus' current interests focus on educational techniques aimed at creating greater awareness and thereby bringing grey literature to the forefront in the medical community.
Email: mmvaska@ucalgary.ca

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

Examples

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

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.

https://doi.org/10.26069/greynet-2026-000.555-tgj
TGJ Volume 7, Number 2 2011

An International Journal on Grey Literature

'System Approaches to Grey Literature'

Summer 2011 – TGJ Volume 7, Number 2

Integration of an Automatic Indexing System within the Document Flow of a Grey Literature Repository 65
Jindřich Mynarz and Ctibor Škuta (Czech Republic)

An Analysis of Current Grey Literature Document Typology 72
Petra Pejšová (Czech Republic) and Marcus Vaska (Canada)

A Profile of Italian Working Papers in RePEc 81
Rosa Di Cesare, Daniela Luzi, Marta Ricci and Roberta Ruggieri (Italy)

From OpenSIGLE to OpenGrey : Changes and Continuity 93
Christiane Stock and Nathalie Henrot (France)

GL Transparency: Through a Glass, Clearly 99
Keith G. Jeffery (United Kingdom) and Anne Asserson, (Norway)

Invenio: A Modern Digital Library System for Grey Literature 105
Jérôme Caffaro and Samuele Kaplun (Switzerland)

Colophon 62

Editor's Note 64

Advertisement

Refdoc.fr – Institute for Scientific and Technical Information 92
De Gruyter Saur, *Grey Literature in Library and Information Studies* 98
J-STAGE – Japan Science and Technology Agency 113

On the News Front

New Password Protected Access to LIS Serials 109
Thirteenth International Conference on Grey Literature – Conference Program 110
The Grey Circuit, From Social Networking to Wealth Creation – GL13 Registration 112

About the Authors 114

Notes for Contributors 115