https://doi.org/10.26069/greynet-2026-000.608-tgj
ISSN 1574-1796

An International Journal on Grey Literature

Autumn 2007 – TGJ Volume 3, Number 3

'Mapping Grey Resources'

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

https://doi.org/10.26069/greynet-2026-000.608-tgj

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 3, Number 3, autumn 2007. - *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:
€100 individual, including postage & handling
€215 institutional, including postage & handling

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

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

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

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

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

About GreyNet, 1992-2007
The Grey Literature Network Service celebrates this year its 15th Anniversary. GreyNet 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. Currently, GreyNet is engaged in the development of a distance learning course for graduate and post-graduate students.

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.

https://doi.org/10.26069/greynet-2026-000.608-tgj

Contents
"MAPPING GREY RESOURCES"

https://doi.org/10.26069/greynet-2026-000.608-tgj

Find the Piece That Fits Your Puzzle ...

The Grey Literature Report

from

The New York Academy of Medicine

Focused on health services research and selected public health topics, the Report delivers content from over 250 non-commercial publishers on a bi-monthly basis.

Report resources are selected and indexed by information professionals, and are searchable through the Academy Library's online catalog.

Let us help you put it all together; subscribe to the Grey Literature Report today!

The New York Academy of Medicine

For more information visit our website: www.greylit.org

Or contact us at: greylithelp@nyam.org

https://doi.org/10.26069/greynet-2026-000.608-tgj

EDITOR'S NOTE

In this issue of TGJ, a selection of articles originating in the GL-Conference Series is brought together under the title 'Mapping Grey Resources'. The first three articles deal with the impact of grey literature on water management in coastal and aquatic environments. The four articles that follow, deal with requirements for accessing grey literature in electronic environments. Here, the key issues of repositories and metadata are discussed from a variety of subject related fields.

One of the articles in this issue is by a team of authors who are recipients of the GreyNet Award 2007. They are Todd A. Chavez, Anna H. Perrault, and Pete Reehling from the University of South Florida (USF) and Courtney Crummett from the National Library of Medicine (NLM). Their work on the impact of grey literature in advancing global karst research is an example of excellence. Professor Perrault will receive the award on behalf of the team during GreyNet's Annual Award Dinner held this year on December 9, 2007 in Antwerp, Belgium. Professor Perrault will also present during the Ninth International Conference on Grey Literature (GL9) a follow-up to their research project.

To a great extent, developments in grey literature over the past couple of decades are due to the work of such researchers and their respective organizations. For, it is they who have provided solid foundations in this field of information upon which to construct and test theory as well as its application.

Under the auspices of the Flemish Government, GL9 will offer a unique opportunity to bring together information professionals from around the world. This two-day S&T conference will provide a mix of plenary and panel sessions, as well as an Information Walk-Thru featuring sponsored stands, poster presentations, and P&S reviews. GreyNet will also take the opportunity to exhibit, showcasing 'The Grey Journal' and other of its products and services.

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

https://doi.org/10.26069/greynet-2026-000.608-tgj

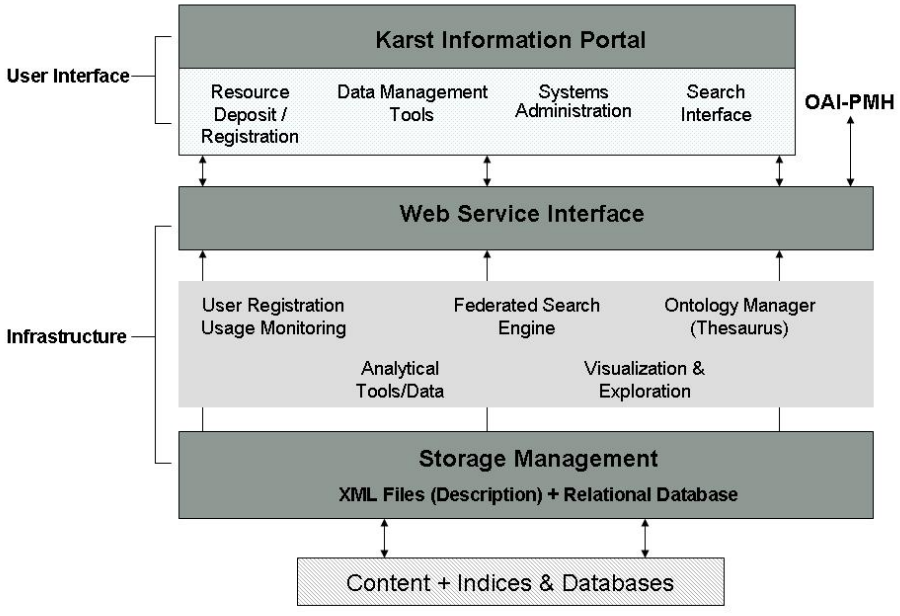
The Impact of Grey Literature in Advancing Global Karst Research: An Information Needs Assessment for a Globally Distributed Interdisciplinary Community *

Todd A. Chavez, Dr. Anna H. Perrault, Pete Reehling, and Courtney Crummett
(United States)

Abstract
A survey of the global karst community was conducted in 2006. The survey was distributed via the World Wide Web to known karst researchers. The instrument was designed to generate an initial inventory of core grey information types, to assess levels of usage of grey information by the respondents, and to gauge the karst community's willingness to participate in building and expanding both this collection and the associated controlled vocabularies.

Background
In 2005, an interdisciplinary work group of faculty, librarians, and graduate students was convened under the auspices of the Dr. Kiran C. Patel Center for Global Solutions at the University of South Florida to discuss global information needs. The group quickly focused upon water issues and then more specifically karst, a very complex and vulnerable type of geologic landform (Drew and Hotzl). Following these deliberations, the group initiated a study to determine the feasibility of constructing a global information portal to be hosted and maintained by the libraries in collaboration with the Patel Center and related academic departments.

In January 2006, a group of 29 scientists, information specialists, and policy makers representing 18 organizations from across the globe met in Carlsbad, New Mexico to explore development of the Karst Information Portal (KIP) to serve as a repository for karst information, to advance collaboration among the international community of karst researchers, and to promote knowledge discovery through innovative applications of metadata. Figure 1 depicts the architecture of the proposed portal.



Adapted from Snow *et al.* "Cybertools and Archaeology." *Science* 311 (17 Feb 2006): 958-9.

* First published in the GL8 Conference Proceedings, February 2007

Figure 1. Graphic Representation of the Karst Information Portal.

The Context

Karst is a globally-distributed terrain resulting from the dissolution of soluble rocks such as limestone and dolomite. This dissolution occurs when rain water infused with carbon dioxide passes through layers of soil and bedrock (see Figure 2). Karst regions contain aquifers and common geological structures such as sinkholes, springs, and caves. The relationship between karst landscapes and water resources evokes the need for greater understanding of the issues underlying these formations.

Figure 2. Karst Terrain (Natural Resources Canada).

The karst research community and its knowledge base are fragmented, globally distributed, highly interdisciplinary, and at the same time, essential to comprehensive understanding of many social, environmental, and health challenges. In a recent study of four widely-used indices covering relevant "white" content, 4,300 individual searches using 632 karst-related terms culled from appropriate thesauri found that, over the period 1960-2005, publication on cave and karst themes has increased substantively. (Florea, Fratesi, and Chavez). Fifteen years earlier, Bichtler (1991) found that grey literature was an important component of the information used by geoscientists. As research into the potential of karst for benefit or hazard to humanity intensifies, information integration and linkages promoting collaboration and connectivity among scientists, decision-makers, educators, and the general public are essential.

Karst researchers are faced with three inextricably related challenges: 1) discovering and evaluating relevant information sources, 2) obtaining and preserving "grey" karst information sources, and 3) providing interdisciplinary linkages among karst scientists to bring about knowledge discovery and communication. In order to construct a portal that contained information and services most useful to karst researchers, a needs assessment was performed.

Global Needs Assessment

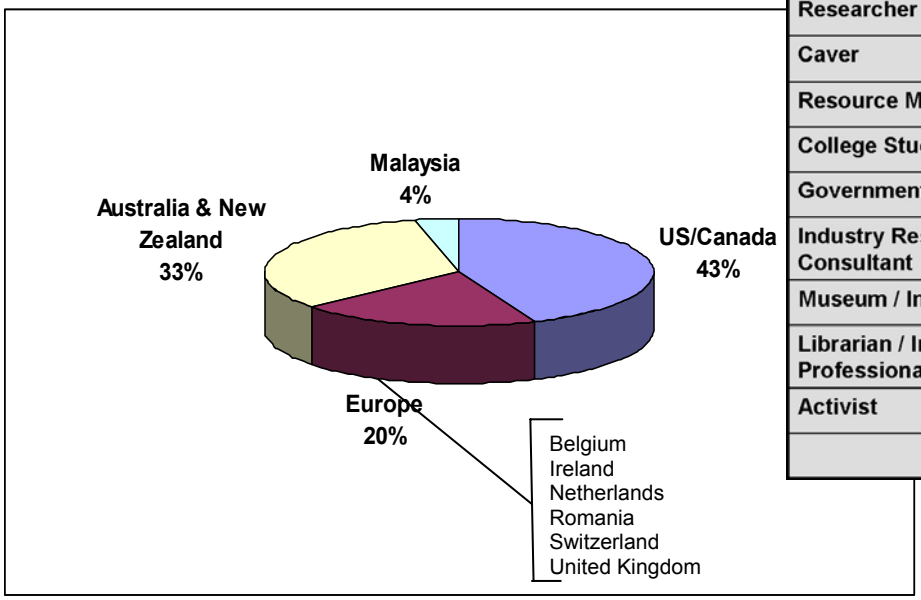
Information specialists from the Libraries and the School of Library and Information Science at the University of South Florida planned and conducted a global information needs assessment for the KIP. The survey was designed to elicit responses in three categories of information need: 1) information content (e.g. format, subjects, and organization); 2) services (e.g. blogs, newsfeeds, and tagging services); and 3) research tools (e.g. data-mining and computational utilities).

Instrument Design

A questionnaire was constructed in Survey Monkey and distributed using a "Snowball" sampling technique through targeted websites to reach the global interdisciplinary karst community and sent via e-mail to a list of karst researchers compiled during the 2006 KIP Planning Workshop. The recipients of the survey were encouraged to respond and forward the survey on to their colleagues and others interested in karst research in some capacity. It was the hope of the survey team that the list of karst researchers would be enlarged through this process and thus increase the data obtained from the survey.

Results of the Survey

A total of 66 responses were received (it should be noted that seven of the 19 survey questions permitted multiple responses and for this reason response totals may exceed 100 percent). Although the number of respondents was not large, the sample is broadly representative by geographic region and occupation and interests (see Figure 3). The countries represented are globally distributed, with the United States and Canada (24) and Australia and New Zealand (17) being the most heavily represented. Responses were also received from six countries in Europe and two responses originated in Malaysia.



Occupation/Affiliation	No.
Researcher	29
Caver	13
Resource Manager	6
College Student	6
Government Agency	4
Industry Researcher / Consultant	3
Museum / Institute	2
Librarian / Information Professional	2
Activist	1
TOTAL	66

Figure 3. Geographic and Occupational Characteristics of Respondents.

The variety of professions or affiliations represented was equally broad, with academic researchers (41 percent), cavers (20 percent), resource managers (10 percent), and students (10 percent) accounting for the majority of respondents. One of the interesting aspects of the distribution of respondents was the proportion who is not academic researchers, cavers in particular who made up one fifth of respondents. One of the findings, not unanticipated, is that karst researchers make up a similar universe as to that of grey literature, that is, they are hard to find and not under the personal equivalency of "bibliographic control."

Locating Karst Information

Respondents were asked to report on "channels" used to locate information about karst-related subjects. For sources of information, the Internet was indicated by 98 percent of respondents. Other sources checked with high frequency were personal correspondence with colleagues (87 percent), conferences and meetings (82 percent), and books (80.6 percent). Of the three types of libraries offered for consideration, personal libraries ranked slightly higher than academic libraries, 84 percent to 81 percent. Public libraries had much lower use at 19 percent (see Figure 4).

Sources of Karst Information	No.
Internet	61
Personal Correspondence w/Colleagues	54
Journal Articles	53
Personal Library	53
Conferences/Meetings	51
Books	50
Academic Library	44
Informal Discussion	42
Public Library	12
Respondents = 62	

Figure 4. Information Channels Used by Survey Respondents.

These findings are consistent with information-seeking behavior research which has shown that people turn to family or friends as their first source of information and only after those close personal sources do they use more formal sources such as libraries (Brown; Given; Hertzum, *et al*; Savolainen). In the specific case of geoscientists, Bichtler (1989) found that personal contacts were extremely important sources of information, particularly when time constraints are significant. Among the two most numerous respondent categories, academic researchers and cavers, personal communications are extremely important, with 100 percent of the cavers and 82.7 percent of the researchers using personal contacts and colleagues as a source if information.

Although their conclusions are limited to a US context, researchers associated with the Pew Internet & American Life Project report that 73 percent of all American adults regularly use the Internet to access information, by far the most common source of information after family and friends (Madden 3). This trend was also observed in the present study.

Context for Grey Literature in Karst Research

Using information derived from interviews conducted during the January 2006 KIP Planning Workshop and with reference to the literature (Grey Literature Network Service; Suloff, *et al*), an extensive list of 46 information types and formats that could be considered grey literature was compiled for the survey. To build a shared understanding of the terminology, respondents were presented with the definition of grey literature adopted during the Third International Conference on Grey Literature:

"[T]hat which is produced by government, academic, business, and industries, both in print and electronic formats, but which is not controlled by commercial publishing interests and where publishing is not the primary activity of the organization." (Farace)

The phrase "non-refereed and self-published documents generated by speleological groups and other non-governmental groups/individuals such as expedition reports," was appended to the core definition to accommodate known grey information types of specific relevance to the karst community.

When asked if the respondents had used grey literature in their work or research, out of 58 responding, 56 (96.6 percent) said they had. The four most commonly used grey information types are conference proceedings/papers, trip and cave reports, theses/dissertations, and maps in any format. The reported incidence of use of conference proceedings and papers is consistent with a 2003 study by Michael Noga in which the researcher found that citation frequencies of proceedings in a selection of geoscience journals indicated that "conference papers in journals are used to the same extent as research journal articles and that some proceedings are used even more"

(19; for a competing assessment of the use of conference papers in published research, see Lacanilao).

Two subsequent questions were designed to identify the types of grey literature 1) used and 2) produced by survey respondents. A third question used the same list of potential grey information types to identify those that respondents found difficult to locate. Table 1 lists, in declining order of frequency of use, the responses to the three questions. For comparison purposes, levels of use are accompanied by the number of responses for grey information source producers and for respondents reporting difficulty locating particular information types.

Table 1. Responses Indicating Use, Production and Difficulty Locating Grey Information Sources.

Information Type (Listed in Order of Frequency of Use)	Responses (No.)		
	Use Type n=58	Produce Type n=52	Difficulty Locating Type n=49
Conference Proceedings and/or papers	50	42	23
Trip and cave reports	49	40	16
Theses/dissertations	47	28	25
Maps (any format)	47	32	14
Non-governmental technical/research reports	44	20	20
Websites with karst/geology content	44	18	4
Government technical/research reports	42	20	21
Images (digital format)	41	34	12
Newsletters	36	24	9
Association and organizational publications and records	34	25	14
Papers and correspondence of researchers	29	22	11
Satellite data	29	3	7
Government documents	28	9	9
Colleague files	28	12	3
Archival materials	28	15	10
Geospatial or GIS data	28	15	9
Modern photograph collections	27	18	8
Cave entrance databases	27	23	18
Files in park service offices	26	19	13
Preprints	25	18	10
Datasets	25	19	9
Historical photograph archives	24	8	11
Translations of foreign-language publications	24	4	15
Monitoring project data	24	15	9
Speeches or invited talks	23	34	3
Working papers (i.e. records of individual or group activities)	23	19	6
Old published karst papers (pre-1923)	23	0	13
Files in government agency offices	22	17	11
Consultant's reports	22	17	12
Management literature	22	13	10
Water-tracing data and databases	22	14	10
Videos (digital or analog)	21	13	3
Cave registries	21	14	12
International Union of Speleology (UIS) abstracting products	19	7	2
Contract reports	18	14	10
Oral histories	18	8	5

Information Type (Listed in Order of Frequency of Use)	Responses (No.)		
	Use Type n=58	Produce Type n=52	Difficulty Locating Type n=49
Catalogs from libraries/collections not linked in any system	16	10	10
Research proposals	15	30	3
Grant applications	14	23	4
Morphology-type databases and collections	13	7	5
Statistical documents or reports	11	7	3
Research permit system information	8	9	3
Databases not linked in any system	8	13	5
Registries of current research initiatives	7	5	7
Realia and artifacts	6	0	1
Audio tapes/files (digital or analog)	4	6	0

Not surprisingly, academic researchers account for a significant percentage (74.1) of respondents who report producing grey information in some format. Roughly 69 percent of the researchers contribute to conference proceedings, deliver speeches/invited talks or generate images while 55 percent produce trip and cave reports and 51.7 percent create or contribute to cave registries or entrance databases. What is surprising is the finding that 84.6 percent of self-identified cavers report producing grey literature, with trip and cave reports and cave registries or entrance databases the most frequent contributions. Responses also indicate that five of the six college or university student respondents produce grey information including conference papers, theses/dissertations, trip and cave reports, images, datasets, and maps.

It is interesting to note the correlation between the most frequently used sources and levels of production. For example, 42 of the 52 respondents to the question concerning production of grey information report contributing to conference proceedings while 40 respondents generate trip and cave reports. Comparing the pattern of responses reveals potential areas of emphasis for portal designers.
Consistent with the literature, trip and cave reports are a significant form of grey information for geoscientists generally including 18 of 29 academic karst researchers and 11 of 13 cavers responding and as such illustrate the importance of studies that focus on specific knowledge domains. Bichtler, Corbett, Derksen, Haner, and Walcott have independently addressed the specific case of the geological field trip book. Produced by local experts to support excursions into specific field locations, trip and cave reports typically include coverage of transportation resources and relate information about local cultural, geological and geographic features and conditions at a specific point in time (Bichtler, 41-42, 1991). Both grey and "white" publications often contain citations to trip and cave reports or field books, but, because they are often published by organizations lacking an infrastructure to facilitate wide distribution, librarians are hard pressed to acquire copies and once in hand, cataloging is a challenge (Haner 166-7; Walcott). It is only through efforts such as The Stanford Geological Survey Map and Field Notebook Project and the Cornell Laboratory of Ornithology's Science Knowledge and Education Network (SKEN) – and potentially the KIP – that critical domain-specific grey information is likely to be systematically accessible and preserved (Derksen; Kelling, *et al*).

During the January 2006 KIP Planning Workshop, discussions with participants revealed the importance of visual information in geoscience research. In their own right, images, videos, and maps are key sources of information about geological subjects, and they are often integrated into other information types including trip and cave reports, theses/dissertations, and websites. Responses from the four most numerous respondent categories supported this anecdotal conclusion (see Figure 5), and coupled with the fact that these formats pose unique challenges for librarians seeking to exert bibliographic control and thereby enhance information discovery and access, suggest a potential content niche for portal planners.

	Images		Maps	
	% Use	% Produce	% Use	% Produce
Academic Researchers	68.9	68.9	75.8	55.5
Cavers	46.0	53.8	76.9	61.5
Students	50.0	33.3	83.0	33.3
Resource Managers	83.3	33.3	66.6	0.0

Figure 5. Use and Production of Images and Maps.

Respondents reported difficulty locating all of the grey information types presented except audio tapes/files. The positive correlation between use and difficulty in locating grey information is relevant in many information type categories but fails markedly in the case of websites with karst/geology content – many respondents report using them (n=44), but only four indicated that they were encountering difficulties. This bodes well for the directions outlined by KIP planners in that, if properly designed and maintained, the site can have a significant positive impact for karst researchers seeking related information, regardless of format. The number of respondents encountering difficulties in locating theses/dissertations, association publications and records, and pre-1923 karst papers suggests that digitization is an important service dimension, a finding confirmed by responses to subsequent questions.

It was somewhat surprising to learn that relatively few respondents reported difficulty locating karst-related geospatial or GIS data. Just over 43 percent of respondents use GIS data and 28.8 percent report producing this information type, but only 18.7 percent encounter difficulties locating needed geospatial data with academic researchers reporting the lowest level of difficulty (17.2 percent) and students the highest (33.3 percent). Possible explanations for this finding include the substantial efforts of the GIS community to ensure that metadata standards are maintained and systematically applied and the proliferation of geospatial portals on the WWW. Because geospatial/GIS data are by definition born digital, dissemination via the WWW is greatly facilitated.

Finally, several questions presented an array of potential activities, services, and capabilities that are being considered for inclusion in the KIP and asked respondents to rank their importance with the context of their research interests. Embedded within these questions were additional queries about grey literature. In all instances, respondents considered treatment of grey information a key function of the portal:

- 99 percent considered grey literature's inclusion very or somewhat important;
- 96 percent rated grey literature digitization very or somewhat important; and
- 85 percent responded that evaluating the authenticity/reliability of karst-related grey literature was an important portal service.

Archiving Grey Literature

Bichtler (49, 1991) describes the important role that geologists play in preserving the grey information that they produce. Survey responses suggest that this dimension of grey information management continues to pose challenges. Eighty-nine percent of the respondents to the survey reported that they produce grey information in some form, but an alarming 28.3 percent do not formally archive their information and just over 75 percent reported using a personal archiving space. This trend is not limited to a single format of grey information – in 2002, Maples, Cutler, and Dickenson warned that the geoscience community must act to preserve data and collections of physical items under a schema that incorporates a standard format for bibliographic citation promoting discovery and access (11).

Within the four most numerous respondent categories, 68.9 percent of the academic researchers reported that they formally archive their grey information in some manner, with resource managers reporting archival efforts in 66 percent of cases. Cavers and college or university students similarly archive grey information (61.5 and 50 percents respectively). One observation is warranted: despite likely ease of access to academic libraries, only 12 of 29 academic researchers use libraries or repositories as archival resources, a finding that suggests potential opportunities for librarians.

Where are you archiving your grey information?	No.
Personal Archiving Space	42
Club/Association Archives	19
Library or Institutional Repository	16
No Formal Archiving	15
Government Files	3
Various Websites	2
Other	3
Respondents = 53	

Would you use the KIP as a personal digital repository?	%
Yes	43.6
Yes, but with conditions	37.4
No	20.0
Respondents = 55	

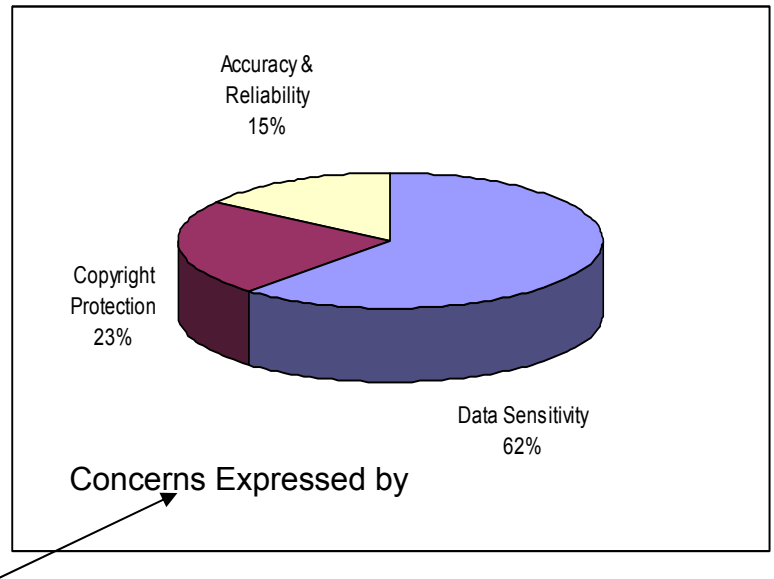


Figure 6. The KIP as Grey Information Repository.

Any archival strategies must take into account an issue not anticipated at the onset of the survey, namely the matter of data sensitivity and the potential for improper use of contributed data in two areas of inquiry: cave entrance locations and water-tracing information. According to unsolicited comments, when asked if they would use the KIP as a personal digital repository, a number of respondents were concerned that cave entrance locations remain non-specific to protect fragile cave ecosystems from potentially damaging visits by non-specialists. Similar concerns for water-tracing data were also expressed: precise descriptions of hydrologic systems could endanger the resources by enabling resource destruction through casual and more nefarious activities.

Concerns for copyright protections, attribution, and information authority/reliability were also expressed, and were anticipated when the survey was designed. Mechanisms for managing these concerns are known and are presently within the expertise of the library and information science community.

Karst Information Portal Services and Capabilities

In addition to the role of the KIP in preserving and providing physical and intellectual access to grey literature, other questions focused on its potential for developing connectivity and promoting collaboration via services or capabilities such as file sharing, RSS feeds, blogs, data management tools, web indexing, and directory services. The select list of potential services or capabilities was developed during the January 2006 KIP Planning Workshop. Figures 7 and 8 summarize responses to these queries.

KIP Activities: Importance to Respondent Research	Very Important	Somewhat Important	Not Important	No Opinion
Digitize Key Grey Literature Resources	37	14	1	1
Develop Potential Karst DBs	34	17	0	2
Evaluate Authenticity & Reliability of Karst-Related Grey Information	24	21	7	1
Identify Global KIP Partners	22	24	6	2
Develop Registry of Karst Programs & Conferences	20	27	6	1
Develop Registry of Karst Community	19	33	3	1
Build Digital Collection of Karst Policy Resources	18	27	9	1
Build Digital Karst Image Collection	14	26	10	3
				n=55

Figure 7. Potential Karst Activities.

Materials to Include in the KIP: Importance to Respondent Research	Very Important	Somewhat Important	Not Important	No Opinion
Key Karst Information Sources	39	16	1	0
Grey Literature	34	21	1	0
Speleological Literature DB	32	16	7	0
Karst Community Contact Information	23	28	3	0
Karst Specific GIS Data	22	22	8	1
Data	19	25	6	3
Karst Image DB	12	26	11	4
Multilingual Dictionary & Controlled Vocabulary (Thesaurus)	9	24	15	4
				Respondents = 55

Figure 8. Potential Portal Content.

Early investigations into the role of grey literature in karst research conducted during the January 2006 KIP Planning Workshop suggested that evaluating the authenticity and reliability of grey information was a controversial portal function. Anecdotal evidence indicated that many workshop participants felt palpable distrust of any attempt to assess levels of authenticity or reliability despite their concerns for these qualities in grey literature generally. This finding was borne out by survey responses, with 28 percent of academic researchers responding to the questioning judging that this function was not an important portal service. In contrast, all of the cavers and four of the five college or university students responding to the query considered evaluation of grey information very or somewhat important.

Although the predominance of interest in services and collections relevant to grey information is readily apparent, significant support for such services as file sharing, analytical tools, and email is documented in the survey results (see Figure 9). Given developments in search engine performance, it is notable that such a large percentage of respondents continue to rank searchable link collections and search tools (83.9 and 60.7 percents respectively) as important for inclusion in the portal. The authors interpret these results as indicators of the continued relevance of domain-specific information portals as mechanisms to promote information discovery in specialized areas.

What types of services or capabilities would make a portal valuable to you?	No.
Searchable Collection of Links to Karst Information	47
Access to and Evaluation of Grey Literature	45
Search Tools	34
File Sharing Services	30
Member Directory	21
Tools for Analyzing Data	21
Capabilities to Export Bibliographic Data	20
Tools for Creating Customized DBs	19
Email Service	18
Pattern Recognition Search Software	15
Wikis	12
RSS Feeds	9
Podcasts	7
Other (journals)	2
Respondents = 56	

Figure 9. Potential Services and Capabilities.

Participation in and Support for the KIP Project

At the conclusion of the survey, respondents were queried regarding potential involvement in portal design, development, and maintenance activities. Fifty-four percent of the respondents indicated that they would prefer participation as a "General User," while 34.5 percent indicated participation preference as a "Minor Contributor." Only six of the 55 respondents to this question indicated willingness to serve as a major contributor or project partner. Given the sampling methodology, it is difficult to interpret the impact of this distribution of responses in the context of project sustainability.

Twenty-eight respondents provided a comprehensive list of potential professional meetings and conferences amenable to marketing and promoting the project. When asked to suggest appropriate target audiences for the KIP, 96.2 percent of the respondents indicated that academic researchers or professors were appropriate, with college or university students (88.7 percent), cavers (83 percent), resource managers (69.8 percent), and museum or institute affiliates (67.9 percent) completing the top five responses. This was an encouraging finding in that it supports the project planners' goal of designing a resource with broad appeal and utility.

Conclusion and Recommendations

In his piece, "The Role of Grey Literature in the Sciences," Professor Irwin Weintraub asserts that,

"In a world in which free trade and instantaneous communication have eliminated many of the barriers to information flow, grey literature is gaining greater importance as a source of information for much of the world's population. It is an indispensable resource for an informed and enlightened public and will undoubtedly continue to serve as a necessary supplement to journal literature well into the future."

This study supported Weintraub's general characterization in the specific case of the interdisciplinary domain of karst studies – the impact of grey information sources on the globally-distributed karst research community is significant and, according to their survey responses, growing. This trend is consistent with developments in similarly interdisciplinary research domains including library and information science (Aina), the health sciences (Alberani, et al; Dunn), marine and fisheries science (Cordes), economics (Mili), and transportation studies (Osif). Based on the consensus definition presented at the beginning of the instrument, 70.9 percent of respondents indicated that grey literature is currently a "very important" source of information for karst researchers, with 14.5 percent stating that it will be more important in the future. Several respondents noted that, like astronomy and ornithology, cave and karst

research benefits from the activities of non-academic individuals and as such would continue to generate (and consume) grey information. Citing recent articles in the journal *Nature*, one respondent made specific mention of the "ongoing rebellion" in the peer-review process and its role in promoting the importance of grey information.

Although these responses are consistent with the prevailing trend of increasing use of grey information in geoscience research generally, challenges persist. Fifteen years ago, Bichtler listed limited distribution, poor bibliographic control, and nonstandard formats as the primary challenges for geoscientists as they accessed grey literature (39-40). In 2003, Mary Scott described the state of bibliographic control of pre-1900 geoscience literature as steadily improving, but not yet fully satisfactory (108). The survey respondents' answers to questions concerning difficulties in locating grey information suggest that Bichtler and Scott's conclusions continue to ring true.

Consider the seemingly innocuous role of grey information citation patterns. In their 1993 analysis of references cited in US Geological Survey publications, Butkovich and Musser found that grey information was frequently incorporated in the text, but was often omitted from bibliographies, a condition that they suggested devalued the material's role in the research process. In her subsequent exploration of this phenomenon in 2002, Musser found that the practice continued and "misleads scholars new to the field, historians of science, and other interested in the preservation of the research resources of the geosciences" (6). Simply improving bibliographic control over grey information and enhancing channels for discovery and access – both primary goals of the KIP initiative – will advance karst research, facilitate informed decision-making, and develop future research agendas, in the process transforming global understanding of karst terrains.

One of the primary purposes of the instrument was to generate an initial inventory of core "grey" information resources as well as gauge the karst community's willingness to participate in building and expanding both this collection and the associated controlled vocabularies. Survey results will assist in formulating guidelines for the collaboration-connectivity requirements of a research community spanning the globe. Future research should focus on 1) developing strategies to address concerns surrounding information security/sensitivity and 2) refining the list of key grey information sources for karst researchers with particular emphasis on those unique to the field.

When implemented, the KIP can serve as a model for similar studies of global interdisciplinary communities and the gathering and synthesis of literature to support the research needs of that community.

References

Aina, L. O. "Grey Literature and Library and Information Studies: A Global Perspective." *International Journal on Grey Literature* 1.3 (2000): 179-82.

Alberani, Vilma, Paola De Castro Pietrangeli, and Anna Maria Rossi Mazza. "The Use of Grey Literature in Health Sciences: A Preliminary Survey." *Bulletin of the Medical Library Association* 78.4 (1990): 358-63.

Bichteler, Julie. "Information-Seeking Behavior of Geoscientists." *Special Libraries* 80.3 (1989): 169-79.

---. "Geologists and Gray Literature: Access, Use, and Problems." *Science & Technology Libraries* 11.3 (1991): 39-50.

Brown, Mary E. "A General Model of Information-Seeking Behavior." *ASIS '91: Proceedings of the 54th ASIS Annual Meeting, October 27-31, 1991, Washington, DC,*. Ed. José-Marie Griffiths. Medford, NJ: American Society -for Information Science, 1991. 9-14.

Butkovich, Nancy J., and Linda R. Musser. "Analysis of References Cited in US Geological Survey Bulletins, Circulars, and Professional Papers." *Finding and Communicating Geoscience Information: Proceedings of the 28th Meeting of the Geoscience Information Society, October 25-28, 1993, Boston, Massachusetts*. Ed. Connie Wick. Alexandria, VA: Geoscience Information Society, 1994. 167-171.

Corbett, Robert G. "Field Trip Guidebooks Need Not Be Gray Literature." *Individual Workstations: Information Supermarkets for Geoscientists*. Ed. Joan V. Lerud. Alexandria, VA: Geoscience Information Society, 1989. 113-22.

Cordes, Ruth. "Is Grey Literature Ever Used? Using Citation Analysis to Measure the Impact of GESAMP, an International Marine Scientific Advisory Body." *The Canadian Journal of Information and Library Science* 28.1 (2004): 49-67.

Derksen, Charlotte R. M, and Julie Sweetkind-Singer. "Accessing and Preserving Field Maps and Notebooks: The Stanford Geological Survey Map and Field Notebook Project." *New Heights in Geoscience Information: Access and Technology: Proceedings of the 37th Meeting of the Geoscience Information Society, October 27-31, 2002, Denver, Colorado, USA*. Ed. Lisa G. Dunn. Alexandria, VA: Geoscience Information Society, 2003. 17-23.

Drew, D., and H. Hotzl, eds. *Karst Hydrogeology and Human Activities: Impact, Consequences, and Implications*. Rotterdam: International Association of Hydrogeologists, 1999.

Dunn, Kathel. "Impact of the Inclusion of Grey Literature on the Scholarly Communication Patterns of an Interdisciplinary Specialty." *Work on Grey in Progress: Sixth International Conference on Grey Literature, New York Academy of Medicine, NYAM Conference Center, December 6-7, 2004, New York, N.Y., USA*. Amsterdam: TextRelease, 2004. 50-54.

Farace, D.J. "Forward." *Perspectives on the Design and Transfer of Scientific and Technical Information: Third International Conference on Grey Literature, Jean Monnet Building, November 13-14, 1997 Luxembourg*. Ed. D.J. Farace. Amsterdam: TransAtlantic GreyNet, 1998.

Florea, Lee, Beth Fratesi, and Todd Chavez. "The Reflection of Karst in the Online Mirror: A Survey within Scientific Databases, 1960-2005." *Journal of Cave & Karst Studies* [Forthcoming].

Given, Lisa M. "The Academic and the Everyday: Investigating the Overlap in Mature Undergraduates' Information-Seeking Behaviors." *Library & Information Science Research* 24.1 (2002): 17-29.

Grey Literature Network Service. "Document Types in Grey Literature." *GreySource: A Selection of Web-Base Resources in Grey Literature*. <http://www.greynet.org/greysourceindex.html> (16 July 2006).

Haner, Barbara E. "Guidebook Citation Patterns in the Geologic Journal Literature: A Comparison between 1985 and 1967." *Frontiers in Geoscience Information: Proceedings of the Twenty-Fourth Meeting of the Geoscience Information Society, November 6-9, 1989, St. Louis, Missouri, USA*. Ed. Mary B. Ansari. Alexandria, VA: Geoscience Information Society, 1990. 159-69.

Hertzum, Morten, Hans H. K. Andersen, Verner Andersen, and Camilla B. Hansen. "Trust in Information Sources: Seeking Information from People, Documents, and Virtual Agents." *Interacting With Computers* 14 (2002): 575-599.

Kelling, S., P Allen, R. Bonney, and A. Poole. "Building a User Base Around Scientific Publications: Editing Online Content and Annotating Scientific Materials." 2004. *Cornell Laboratory of Ornithology*. 2003. <http://www.birds.cornell.edu/it/docs/report_sken.pdf> (8 January 2006).

Lacanilao, Flor. "Continuing Problems with Gray Literature." *Environmental Biology of Fishes* 49 (1997): 1-5.

Madden, Mary. "Internet Penetration and Impact, April 2006." 2005. *Pew Internet & American Life Project*. 2006. <http://www.pewinternet.org/pdfs/PIP_Internet_Impact.pdf> (14 October 2006).

Maples, Christopher G., Paul Cutler, and Tamara Dickenson. "Geoscience Data and Collections Preservation: What Can the Professional Community Do Now?" *New Heights in Geoscience Information: Access and Technology: Proceedings of the 37th Meeting of the Geoscience Information Society, October 27-31, 2002, Denver, Colorado, USA*. Ed. Lisa G. Dunn. Alexandria, VA: Geoscience Information Society, 2003. 11.

Mili, Fethy. "Trends in Publishing Academic Grey Literature: Examples from Economics." *International Journal on Grey Literature* 1.4 (2000): 157-66.

Musser, Linda R. "What Geoscientists Cite: A Comparative Study." *New Heights in Geoscience Information: Access and Technology: Proceedings of the 37th Meeting of the Geoscience Information Society, October 27-31, 2002, Denver, Colorado, USA*. Ed. Lisa G. Dunn. Alexandria, VA: Geoscience Information Society, 2003. 5-10.

Natural Resources Canada. "What is Karst?" 2005. *Geoscape Nanaimo: Geoscience for Central Vancouver Island Communities*. 27 October 2005. <http://geoscape.nrcan.gc.ca/nanaimo/karst_e.php> (6 June 2006).

Noga, Michael Mark. "Conference Proceedings in Geoscience Journals: What's the Use?" *Geoscience Information Horizons: Challenges, Choices, and Decisions: Proceedings of the 38th Meeting of the Geoscience Information Society, November, 2-6, 2003, Seattle, WA, USA*. Ed. Lura E. Joseph. Alexandria, VA: Geoscience Information Society, 2004. 19-35.

Osif, Bonnie A. "International Transportation Literature: An Analysis of Citation Patterns, Availability, and Research Implications to the Transportation Community." *International Journal on Grey Literature* 1.4 (2000): 149-56.

---. "Transportation Information: A Review of Grey Literature by Format, Language, and Availability." *International Journal on Grey Literature* 1.1 (2000): 12-17.

Savolainen, Reijo. "Everyday Life Information Seeking: Approaching Information Seeking in the Context of 'Way of Life'." *Library & Information Science Research* 17.3 (1995): 259-294.

Scott, Mary W. "Status of Bibliographic Control of Pre-1900 Geoscience Literature." *Geoscience Information Horizons: Challenges, Choices, and Decisions: Proceedings of the 38th Meeting of the Geoscience Information Society, November, 2-6, 2003, Seattle, WA, USA*. Ed. Lura E. Joseph. Alexandria, VA: Geoscience Information Society, 2004. 105-08.

Sulouff, Pat, et al. "Learning About Grey Literature by Interviewing Subject Librarians." *C&RL News* 66.7 (2005): 510-5.

Walcott, Rosalind. "Guidebook Problems from a Librarian's Point of View." *Frontiers in Geoscience Information: Proceedings of the Twenty-Fourth Meeting of the Geoscience Information Society, November 6-9, 1989, St. Louis, Missouri, USA*. Ed. Mary B. Ansari. Alexandria, VA: Geoscience Information Society, 1990. 185-92.

Weintraub, Irwin. "The Role of Grey Literature in the Sciences". 2000. *Brooklyn College Library*. <http://library.brooklyn.cuny.edu/access/greyliter.htm> (15 July 2006).

https://doi.org/10.26069/greynet-2026-000.608-tgj

Searching down the fisheries information web: An initial reflection*

Janet Webster *(United States)*
Patricia Merrikin and Jean Collins *(Italy)*

Abstract
The complex web of information ranges from pure science to applied techniques to management policies. All are important, yet have varying levels of accessibility and authority. Often the apex of the peer-reviewed journal article is considered the most important level discouraging scientists and managers, especially those in developing countries, from populating the other layers of the web that may be considered grey. We are particularly interested in how information produced in developing countries fits into this information web. To do so, we focused on one topic, mangroves, using it as a means to exemplify the complexity of the web of information and reveal publication patterns. Using search tools, we identified 2000-2005 mangrove-related publications regarding Africa and India, analyzed them by author's location and affiliation, publication format and availability. Our preliminary analysis suggests that grey literature remains difficult to discover, the peer-reviewed journal is the most popular means of scientific communication, and open archives are just starting to have an impact on scholarly communication. The web of information would be strengthened by improving the discovery of the grey literature through greater visibility and accessibility.

Introduction
Fishing down the marine food web is an influential concept proposed by Pauly et al. (Pauly and others 1998)to describe the pattern where high trophic level fish species are over-exploited thus disturbing the ecology and economics of the system. Essington et al. (Essington and others 2006) suggests that fishing *through* the food web may better describe the situation as more trophic levels are exploited leading to a greater complexity in fisheries management than simply over-fishing. We contend that the two concepts have analogies in the complex web of information that ranges from pure science to applied techniques to management policies. All are important, yet have varying levels of accessibility and authority. Often the apex of the peer-reviewed journal article is considered the most important level discouraging scientists and managers, especially those in developing countries, from populating the other layers of the web that may be considered grey.

Grey literature tends to be at the bottom of the information web in part because it is hard to identify and access, particularly in the increasingly digital environment. It also sinks because it is not recognized as valuable given peer pressure on people to publish in the commercial journals. Consequently, relevant information is easily overlooked and underutilized especially if it originates in a developing country. This phenomenon may devalue the work of scientists and managers in developing countries, slowing responsible management of the natural environment and its valuable resources. If information is difficult for the end user to find or access, it tends not to be used as readily (Jansen and others 2000). For those involved with collecting or monitoring grey literature, this issue is familiar. Even so, increased documentation is useful for building the case for better indexing and improved discovery tools.

We initiated the following study to test our thesis that the grey literature of fisheries needs to be more visible so it can be more viable in the information web. Researchers and managers use discovery tools that are familiar and readily available. Some of these tools are well established with sophisticated indexing and defined sources (e.g. *CAB* and *Biosis*). Others are emerging as useful tools (e.g. *GoogleScholar*). Comparing results produced using a range of current tools is one way to document what information is visible and accessible, who creates it and where it is published. It also allows some consideration of trends in coverage, although this would take more longitudinal work. This progress report on our study suggests ways to improve visibility and accessibility of the fisheries grey literature.

* First published in the GL8 Conference Proceedings, February 2007

Approach and Methodology

In the study, we focused on information produced on mangroves in Africa and India. Mangroves are productive ecosystems for a variety of natural resources including fisheries. We originally thought to narrow our topic further by limiting mangrove information to that addressing fisheries or aquaculture, but this resulted in very small or null sets in most of the indexes used. Africa and India were chosen as geographic limiters for several reasons: they differ in their development of a scientific community; they are working on developing responsible fisheries management; and mangroves are significant habitats in each area. We also limited our searches to the time period of 2000-2005 in part to keep the results manageable, but also to see if the open access movement was having an impact on access to the literature.

Table 1 lists the search tools and the basic search strategies along with the resulting hits. From these, we chose to analyze the ones in bold. Search strategies varied depending on the capability of the search tool. *GoogleScholar* does not accommodate true Boolean logic or field searching, so the strategy was quite basic. In most of the others, we used field limiters to get more precise results. When possible, we limited to title/abstract/keyword, used geographic limiters or expanded through the thesaurus to better refine the results. On the surface, this appears to skew the overall results as the database results are more precise while *GoogleScholar* less so consequently producing more hits. To remedy this somewhat, we searched *GoogleScholar* by year (a limiter that it provides) and took the first 28 to 30 hits from each year searched to build a set comparable to the *Biosis* and *CAB* sets. This is not a perfect comparison, but gives a solid set of results for analysis.

Table 1:

Search Tool	Search strategy	Total hits	Total hits analyzed
AGRIS (OVID)	MANGROVE* and AFRICA	17	
	MANGROVE* and INDIA	5	
Aquatic Biology, Aquaculture and Fisheries Resources	MANGROVE* and AFRICA	590	
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)	91	74
	MANGROVE* and INDIA	336	
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)	116	112
Biosis (OVID)	MANGROVE* and AFRICA	186	186
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)	37	
	MANGROVE* and INDIA	179	179
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)	24	
CAB (OVID)	MANGROVE* and AFRICA	146	146
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)		
	MANGROVE* and INDIA	200	199
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)		
Food and Agriculture Organization of the UN library catalog	MANGROVE* and AFRICA	2	
	MANGROVE* and INDIA	13	
GoogleScholar	MANGROVE* and AFRICA	3280	176
	MANGROVE* and AFRICA and FISH	2050	
	MANGROVE* and AFRICA and AQUACULTURE	619	
	MANGROVE* and INDIA	2550	179
	MANGROVE* and INDIA and FISH	1350	
	MANGROVE* and INDIA and AQUACULTURE	473	
Horizon	MANGROVE*	30	
OAIster	MANGROVE* and AFRICA	19	
	MANGROVE* and INDIA	3	
Scopus	MANGROVE* and AFRICA	122	100
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)		
	MANGROVE* and INDIA	161	100
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)		
TROPAG (OVID)	MANGROVE* and AFRICA	22	
	MANGROVE* and INDIA	19	

Each result was coded for the following:

- Location of first author
- Affiliation (university, research institute, commercial, government, other)
- Format (journal, book, conference proceeding, report, newsletter)
- Location of publisher
- Available via the Internet
- Publicly accessible in electronic format

Results were analyzed using Excel. All analyzed search results were compiled into a bibliographic database using ProCite. This enabled us to check for duplication among the search results.

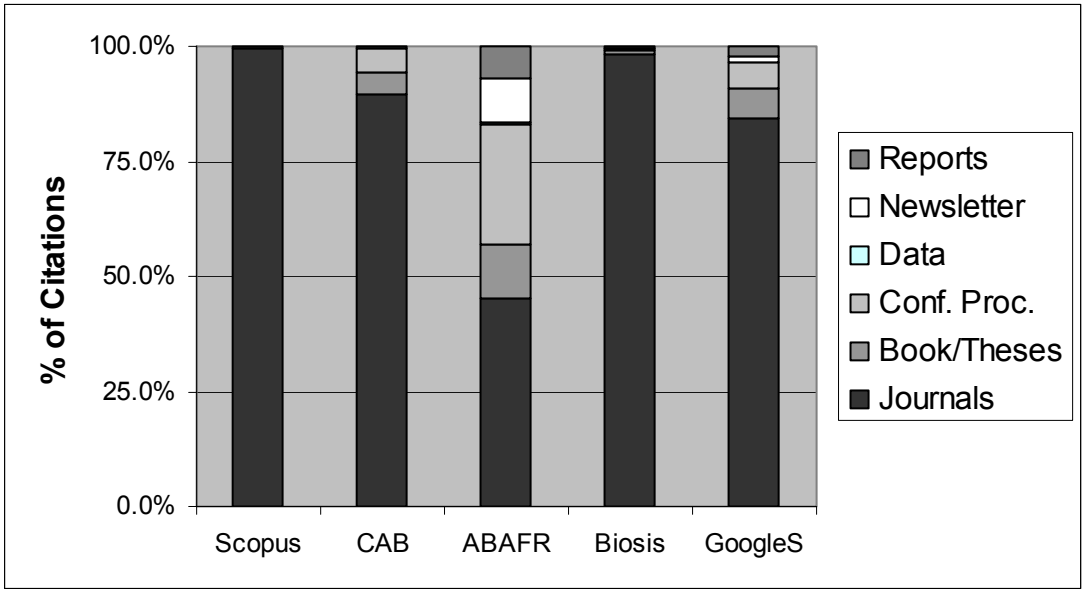
The following presents initial data from the analysis of all results by search tool. We also analyzed the data with duplicates removed, but the findings were similar. The results allow us to examine the efficacy and focus of various search tools. They also give a picture of who is publishing and where. Our primary goal was to discover what mangrove related information can be discovered by an interested user, who is writing it, where is it published, and is it accessible.

Results

What is visible?

Table 1 presents raw data on where material is indexed, but it is difficult to truly compare apples and oranges, *Biosis* to *GoogleScholar*. Figure 1 gives an overview of what is visible using each tool. In general, we can say that the index, *Aquatic Biology, Aquaculture and Fisheries Resources* (*ABAFR*), is the best access tool for the topic. This is not unexpected as its producers draw from the relevant portions of *Aquatic Science and Fisheries Abstracts* and enhance it with appropriate records from a variety of sources including grey literature. It is surprising that databases such as *AGRIS* did not produce more results.

Figure 1: Type of material by search tool



Who is writing?

Comparing India to Africa reveals that while 82.5% of the items on India are authored by scientists in India, only 47.2% of the items on Africa are authored by scientists in Africa (Table 2 &3). Further analysis of second and third authors would be useful as we noticed that many papers did have authors located in Africa, just not as the first author. Also, student authors were assigned to their institution's country and this too does not always reflect nationality.

Table 2: First author's location – India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	1		1			2	0.3%
Asia	2	3		5	4	14	1.8%
Oceania	3	1	1	3	3	11	1.4%
Europe	14	15	9	17	26	81	10.6%
India	76	177	96	149	135	633	82.6%
North/Central America	3	2		5	10	20	2.6%
South America	0			0	1	1	0.1%
Unknown	1		3			4	0.5%
	100	198	110	179	179	766	

Table 3: First author's location – Africa

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	51	85	52	82	52	322	47.2%
Asia		3	1	2	7	13	1.9%
Oceania		2	2	4	8	16	2.3%
Europe	44	46	18	82	80	270	39.6%
India		1		4	3	8	1.2%
North/Central America	3	3	1	11	26	44	6.5%
South America				1	0	1	0.1%
Unknown	2	6				8	1.2%
	100	146	74	186	176	682	

Authors are primarily affiliated with either a university or a research institute (Table 4 & 5). There is a tradition of research institutes in India that is not replicated in Africa (Krishna 1997).

Table 4: Author affiliation - India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Commercial		2	6	3	5	16	1.9%
Government	4	8	6	11	2	31	3.7%
Other	2	20	3	2	1	28	3.4%
Research Institute	33	59	40	68	68	268	32.3%
University	60	105	53	95	103	416	50.1%
Unknown	1	5	2			8	1.0%
	100	199	110	179	179	767	

Table 5: Author affiliation - Africa

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Commercial		3	3	3		9	1.3%
Government	1	2	6	3	5	17	2.5%
Other	2	4	5		7	18	2.6%
Research Institute	17	37	26	59	36	175	25.7%
University	76	91	33	121	128	449	65.8%
Unknown	4	9	1			14	2.1%
	100	146	74	186	176	682	

Scientific publishing also appears stronger in India than Africa (Tables 6 & 7). European publishing venues appear to dominate reflecting the consolidation of scientific publishing. Over time, this trend may diminish in both Africa and India as regional journals gain stature. Also, if commitment to open access continues, then some of the Indian and African journals are well-poised to gain in importance.

Table 6: Location of publication – India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	0	0	0	0	0	0	0.0%
Oceania	4	8	4	9	10	35	4.6%
Australia	2	2	0	0	2	6	0.8%
Europe	33	66	21	74	81	275	35.9%
India	35	98	82	82	61	358	46.7%
North/Central America	26	21	3	14	22	86	11.2%
South America	0	0	0	0	0	0	0.0%
Unknown	0	4	0	0	3	7	0.9%
	100	199	110	179	179	767	

Table 7: Location of publication – Africa

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	15	24	38	25	43	145	21.3%
Asia		1	1	4	2	8	1.2%
Australia	1	2	1	2	2	8	1.2%
Europe	75	99	26	129	100	429	62.9%
India	1	3		7	1	12	1.8%
North/Central America	8	14	1	18	21	62	9.1%
South America						0	0.0%
Unknown		3	7	1	7	18	2.6%
	100	146	74	186	176	682	

The peer-reviewed journal remains the predominant format (Tables 8 & 9). This reflects both the bias of the search tools as well as the bias of academic researchers and their institutions.

Table 8: Format - India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Book/Book Chapter		9	13	2	10	34	4.4%
Conference Proceedings	1	8	29	1	11	50	6.5%
Data		1	1			2	0.3%
Journals	99	181	47	176	153	656	85.5%
Newsletter			12		3	15	2.0%
Newspaper					1	1	0.1%
Report			4		1	5	0.7%
Series			4			4	0.5%
	100	199	110	179	179	767	

Table 9: Format - Africa

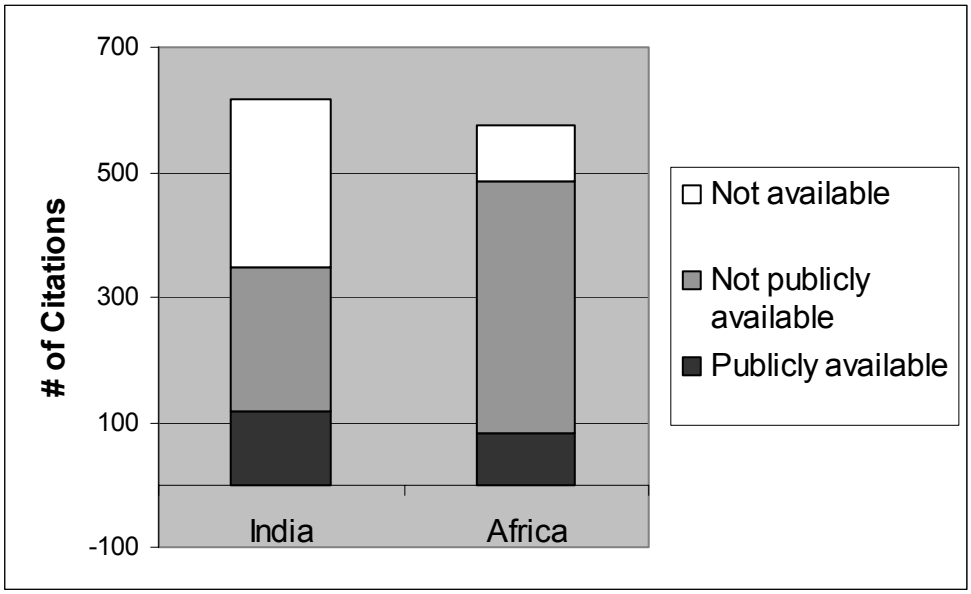
	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Book/Book Chapter		8	9	1	9	27	4.0%
Conference Proceedings		9	19	1	9	38	5.6%
Data		1				1	0.1%
Journals	100	128	36	183	147	594	87.1%
Newsletter			5		1	6	0.9%
Newspaper							0.0%
Report			5		7	12	1.8%
Series				1		1	0.1%
Theses					3	3	0.4%
	100	146	74	186	176	682	

Most material is now available electronically, but little is truly publicly accessible (Figures 2). There is a greater open access in India (30.8% of documents versus 11.8% of African related material). *Current Science*, a journal published by the Indian

Academy of Sciences, is publicly accessible on the Internet back to volume 1, 1932. This is an excellent example as it is also indexed with variability; of the 27 references found, *Scopus* had 0, *CAB* 17, *ABAFR* 4, *Biosis* 17 and *GoogleScholar* 17. The longevity of the journal suggests why *CAB* and *Biosis* index it well, and its open availability is why *GoogleScholar* brings in significant results. Additionally, other Indian journals are trending towards open access. More investigation as to the motivation would be interesting.

African Journals Online (AJOL) provides limited open access so those articles published in journals that are part of AJOL were not counted as publicly available. Researchers in low income countries may request articles for free. There is a sliding scale for others. This portal to African journals is one example of a mechanism to promote African science through both discovery and document delivery.

Figure 2: Electronic accessibility of all mangrove material



Discussion

Quality of results in terms of grey literature

We did not anticipate that all discovery tools would include much that is grey. Of course, there are definitions of grey to consider as well. From our perspective, journals published in developing countries are not grey as they are peer reviewed and distributed appropriately. We were looking for items such as conference proceedings, workshop reports, management studies and government report. We recognized that the type of coverage in commercial databases depends on the subject – pure sciences coverage seem to be almost completely journal articles (e.g. *Biosis*), while applied sciences produce more grey literature (*ABAFR*). Some cover both the pure and applied sciences (*Scopus*, *CAB* and *GoogleScholar*).

Of the tools analyzed, *ABAFR* alone appears to do a decent job of discovering the grey literature. As *GoogleScholar* harvests OAI repositories, we thought it might gather more results. The lack of diverse formats may be due to sparse holdings in OAI repositories which may be remedied as repositories grow in popularity. Another reason may be *GoogleScholar*'s relevancy ranking, and this would need more analysis of all results to see if grey literature sinks to the bottom pages of search results.

Quantity of results

Discovery tools that we expected to be productive were disappointing. For example, *Agris* did not perform well. This was surprising as it was founded to be an international, cooperative information system for the agricultural sciences and technology that utilizes 240 national, international and intergovernmental centers as contributors. In our study, 12 of the 16 hits for mangroves and Africa were from commercial journals, and the remaining four were theses from Norway and a Swedish working paper. A similar search for India produced only five hits all from an Indian journal and *Unasylva* (FAO). Further work could be done to see if *Agris* began with

deeper coverage of non-traditional information, and over time transitioned toward coverage of the journal literature.

GoogleScholar appears to search more broadly with overwhelming search results. Yet, it is still not clear what it is searching. So, while the numbers of hits are impressive, the number of irrelevant hits are also high as it does not discriminate between information items where the search terms are found in the title and abstract from those where the terms appears in the references if listed. It is limited in ways to refine search results. Duplications are a problem as it does not sort out a citation record from the record returned from a publisher's database.

We attempted to use *Scirus* as it claims to be the "most comprehensive science-specific search engine on the Internet" and promises to find results that the other search engines miss (Elsevier Ltd 2006). The search interface is more limited that *GoogleScholar*, so searches return too much with many false hits or too little if title limits are used. There is potential for improving the tool if searching could be limited to abstracts or other elements found in repositories.

Another way to look at quantity of results is examining the overlap among the results from the tools. A preliminary look reaffirms that *ABAFR* returned the most unique results with the only African duplicates being seven shared results with *GoogleScholar* (less than 10% of the 74 African *ABAFR* results) and less than 6% for the Indian results. Other African duplication between each tool and *GoogleScholar* were 18% for *CAB* and *Biosis* and 35% for *Scopus*. A similar pattern was found in the Indian duplicates (*Biosis* 21%, *CAB* 25% and *Scopus* 29%). While more analysis remains, it appears that there is less overlap than was anticipated indicating that each discovery tool has unique qualities, and none are comprehensive.

Access to results

Access to electronic instances of the information reflected the bias of the discovery tool for the most part: most link to the publisher's resource rather than a repository copy. The exception is *GoogleScholar* that did return a fair number of repository copies. It also revealed several electronic theses that did not appear elsewhere. *OAIster*, another free search tool that provides access to open access archives, was disappointing in the quantity of results, but all were relevant.

Access can be deceptive depending on who is doing the searching and where they located. Many believe more is publicly available as it may be for them given institutional subscriptions. Access is highly variable throughout the developing world given services such as *AGORA* that facilitates electronic access to journals from major commercial publishers at little or no cost to low income countries. Most investigation of open access is needed so we can better assess the impact of repositories and open access journals on the information web.

Language of results

We found very few non-English articles. Those few were French, German or Russian. This may be a limitation of where we searched, but probably reflects a trend towards English as the accepted language of science. That said, we remain convinced that there is significant information in non-English languages that are not indexed due to availability, the language and the limitations of search engines. An area of further research would be to explore the modes of discovery for Arabic and Asian language material.

Conclusion

As librarians, we can collect information consistently and direct users to it in a coherent fashion if we understand what is produced and how it is indexed or discovered. We rely on our discovery tools to help us to this. However, as this preliminary study shows, these tools vary in what they find and index. For the most part, the grey literature is missing.

We suggest that Pauly's concept of declining high trophic levels and Essington's concept of greater complexity are useful in describing the web of information. Different tools influence how people exploit the web of information with some seeking out primarily the peer-reviewed, Western published material while others discover more layers of useful information. Separating the so-called grey literature from the peer-reviewed through the tools we use to discover and access information is harmful to the environment. Greater indexing of the grey literature needs to be integrated into existing discovery tools. Given the current state of discovery, we suggest that increased use of open archive repositories has potential for increasing access to grey information if they are constructed to work with discovery tools. Also, working with developers of the newer tools such as *GoogleScholar* and *Scopus* may help improve

searching options and eventual discovery of more grey literature (Powell 2006). The web of information would be strengthened by improving both the visibility and accessibility of lower trophic levels of information or what many of us value as the grey literature of the field.

References

- Elsevier Ltd. 2006. Scirus - About us [Web Page]. Located at: http://www.scirus.com/srsapp/aboutus/. Accessed 2006 Nov.
- Essington T, Beaudreau A, Wiesenmann J. 2006. Fishing through marine food webs. PNAS 103(9):3171-5.
- Jansen BJ, Spink A, Saracevic T. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. Information Processing and Management 36(2000):207-27.
- Krishna VV. 1997. A portrait of the scientific community in India: historical growth and contemporary problems. in: Gaillard J, Krishna VV, Waast R. Scientific communities in the developing world. New Delhi, India: Sage Publications.
- Pauly D, Christensen V, Dalsgaard J, Froese R, Torres FJ. 1998. Fishing down marine food webs. Science 279(5352):860-3.
- Powell A. 2006. Pushing an OpenDOAR [Web Page]. Located at: http://efoundations.typepad.com/efoundations/2006/10/pushing_an_open.html. Accessed 2006 Nov.

https://doi.org/10.26069/greynet-2026-000.608-tgj

Louisiana Coastal Wetlands and Louisiana Coastal Grey Literature: Vanishing Treasures*

Gina R. Costello *(United States)*

Abstract
Over the last century Louisiana has lost an alarming amount of coastal wetlands to coastal erosion. Natural disasters and manmade solutions to problems alike have contributed to this national tragedy. A vast amount of grey literature documenting the history of land loss in Louisiana has been produced, but never collocated for researchers' use. The Louisiana Coastal Grey Literature Project endeavored to locate, organize and provide access to these valuable hidden treasures.

Introduction
Since 1932, when the Department of Natural Resources began keeping thorough, accurate records, Louisiana has lost over 1,900 square miles or 1.2 million acres of coastal land due to coastal erosion. This translates into an area roughly the size of the state of Delaware. The current rate of land loss in Louisiana is a little over 10 square miles per year[1]. These staggering statistics are important to Louisianans and to every U.S. citizen because of the impact this land loss and its subsequent repercussions have on the region and the country.
These statistics lead the Louisiana Governor's Office of Coastal Activities (GOCA) to seek assistance in preserving the important data compiled about Louisiana's land loss over the past 70 years. In October 2004 GOCA and Louisiana State University (LSU) Libraries began an ambitious project to identify, collect, preserve, and provide access to grey literature research relating to the state's battle against coastal land loss. The goal of the Louisiana Coastal Grey Literature Project was to promote the availability of this information as an environmental awareness resource for state and federal agencies, university researchers and the public.
This paper offers a brief background history of the land loss in Louisiana and its causes, information describing the Louisiana Coastal Grey Literature project and its purpose, and a description of the methods that were used to identify grey literature for the project database. Information about the survey designed to assess the needs of potential database users, the creation of the database, and decisions about metadata and organization follow. In conclusion the author offers a summary of the lessons learned and what work still remains to be completed.

Historical Background
Louisiana, which is located in the southern U.S., has historically been a state that is vulnerable to the warm hurricane-prone waters of the Gulf of Mexico and the powerful, capricious flow of the Mississippi River. Early settlers to the land were attracted to coastal and river areas because of the wealth of resources available in and near the water. To live safely in these regions, though, people had to work to control river waters that threatened to flood their land or periodically alter course and change the landscape. Although containment of water flowing in and around Louisiana was necessary for individuals to be able to live and work in the area, it has forever altered the geography of the region and lead to a national crisis that could mean the end for an already fragile ecosystem.
The mighty Mississippi River, which extends from Minnesota to Louisiana where it spills into the Gulf of Mexico, became a transportation and trade route in 1705 when the first boat of cargo travelled down the river[2]. Many people settled in the regions around the Mississippi River and cities such as St. Louis, Missouri, Memphis, Tennessee, and New Orleans, Louisiana sprang up.
Tragic floods that devastated these highly populated areas required the construction of levee systems along the banks of the Mississippi[3]. These levees, which are manmade earthen mounds reinforced by other materials and located on both sides of the river, have saved countless lives and contained a dangerous river that became an important north to south navigation route. However, taming the river has come at a great cost to

* First published in the GL8 Conference Proceedings, February 2007

the wetlands and coastal regions of this area. Naturally occurring wetlands that are home to hundreds of species of birds and wildlife were sustained by the sediment brought by flood waters of the Mississippi.

Manmade levees were not the only cause for the deterioration of this area; the Gulf of Mexico offshore oil and gas industries, which rank numbers one and two respectively in U.S. production[4], brought many changes that drastically affected the landscape. In the early twentieth century the growing oil industry sought better access to Louisiana coastal waters. Manmade canals were dredged in the wetlands abutting the Gulf so barges full of equipment could pass easily to the open water. These canals disrupted the natural flow of water through the wetlands and starved many areas of necessary nutrients.

The changed landscape also affected land and sea life in this area. Saltwater intrusion from the Gulf invaded freshwater areas, killing plant life and forcing both animals and people to move to other locations. Hurricanes such as Betsy, Alison, Katrina and Rita ravaged barrier islands that protected the coast and brought severe wind and water damage to the already vulnerable coastal region. Non-native plants and wildlife that were introduced into the area further disrupted the ecosystem[5].

Before the 2005 hurricanes, Louisiana was losing approximately 25 square miles of land annually due to coastal erosion. One year prior to the 2005 hurricane season U.S. federal agencies had estimated that land-loss in the hurricane-affected areas would be approximately 60 square miles by 2050. However, in just two days in the fall of 2005 Hurricanes Katrina and Rita destroyed 80 square miles of Louisiana coast[6].

Project Description

The Louisiana Coastal Grey Literature project was based on the premise that grey literature about Louisiana coastal wetlands was often difficult to find in many libraries, online databases, or on the web, and that the database end product would rectify this problem. The project staff, which included a project librarian and a graduate assistant who were directed by the Associate Dean for LSU Libraries Special Collections division and advised by a representative from GOCA, set out to locate and collect grey literature from a variety of agencies and organizations. The identified resources would be entered into a relational database designed for the project. An effort to organize coastal wetlands grey literature materials and offer researchers a singular resource had not before been attempted. This project offered a unique opportunity to preserve a wide variety of resources and to increase their value by making them more readily available.

For the purposes of this Louisiana-specific project grey literature was defined, but not limited to, working papers, reports, newsletters, conference proceedings, committee reports, theses and dissertations, and government sponsored research reports. This material, which occurs in print and electronic formats, might include research summaries, facts, statistics, and other data that offer a more comprehensive view of coastal wetland loss in Louisiana.

Locating Resources

Over the past 70 years Louisiana government agencies and private groups have produced a substantial amount of information that examines wetland loss in the state and region. These numerous scientific and technical reports fall into the category of grey literature. The documents are usually highly specialized and quickly released to ensure that researchers have the most current information. However, the information often does not reach far beyond the specific government agency that sponsored the research.

The project staff first developed a list of state agencies, federal agencies and independent organizations that produce or have produced material relating to coastal wetland loss in Louisiana. GOCA provided contact information for individuals at state agencies who might have access to the agency "library" or document resource center, and asked that we concentrate our efforts first on these agencies. Baton Rouge, which is the home of the main campus of Louisiana State University, is also the capitol of the state, so all of the state agencies were within close proximity.

GOCA asserted that there were hidden resources in the various state agency buildings, so the project staff set out to locate these. Many of the agencies had libraries that were no longer staffed and had not been actively collecting materials for many years. The libraries were mostly being used as storage rooms or areas where retiring staff discard their bookshelf materials. Some state agencies did still maintain copies of relevant documents, but these were often housed in cramped, dirty, and unorganized rooms with no cataloging or finding aids for the user.

Project staff developed a set of questions to ask the agency contacts to determine whether any grey literature was housed in the building or if the contact was aware of another resource for such material. Although the project staff did not know quite what they might find at the different agencies, they hoped that the contact would be able to at least direct them to available resources.

The first discovery visit to the Louisiana Department of Agriculture and Forestry yielded little information and no actual documents. The agency did not maintain control of any state documents that were not placed in the state library's collection, nor did they keep any documents that might be useful for this project in a long-forgotten closet or file cabinet. After that disappointing discovery visit, the project librarian contacted the Louisiana Department of Transportation and Development (LDOTD). The building which houses this particular department of LDOTD is a corrugated metal structure located near the National Guard base at the Baton Rouge airport. At some point in the past, an agency employee had gathered grey materials that had been collected in the agency and located them together. The project librarian and project graduate assistant spent two days onsite recording metadata for each of the relevant items. Reports produced by the LDOTD Coastal Management Section were of special interest for the project.

The project librarian and graduate assistant spent over a week culling through the large library in the Louisiana Department of Wildlife and Fisheries (LDWF) headquarters. The library had at one time been organized using a cataloging schema specific to the agency and the librarian on staff. Unfortunately, the database that contained the records had not been migrated to current software and had eventually been lost. Although there were many duplicate items and non-coastal wetlands related items in the collection, many of the materials proved to be very valuable. Relevant documents included research reports that focused on how animals in fish in coastal areas had been affected by wetland loss. The project staff also spent time looking at materials in the office of a retired 30-year veteran of LDWF. Regrettably, these items, which included reels of video footage of a disappearing coast, were located in a musty office in a building that is no longer in full use. There were moldy items and quite a few papers with silverfish that could not be saved.

Throughout the year the project staff visited other state agencies including the Louisiana Department of Natural Resources (LDNR) and the Louisiana Department of Environmental Quality (LDEQ). LDNR contains many departments that work directly with coastal issues, so the project staff spent several days meeting with different agency contacts to discuss available grey literature. A small resource center at the LDNR headquarters contained some resources that had not been previously found at other agencies. The items in this collection were placed their by a LDNR employee and did not seem to be in use by the staff.

The library at LDEQ is staffed by two librarians and other paraprofessional staff, so the experience at this agency was unlike any of the others. LDEQ employees are allowed to check out materials from the library, which was not the case at the other agencies. Aside from the materials in the LDEQ library, all of the materials at state agencies were not being maintained or accounted for. In fact, in talking with different people at the agencies, the project staff discovered that many useful grey literature items that had once been located in the library or resource area were now missing – either on an employee's office bookshelf, or long since lost.

Visits to the agencies proved to be beneficial both in making contact with the individuals who would most benefit from the project database and in identifying resources that might be difficult to find anywhere other than the agency. Often times an agency contact would provide the names of other people with information about grey literature resources. Through these contacts the project staff learned that the LSU Sea Grant administration had decided to close their library, which had not been in active use for several years. Rather than discarding a roomful of materials that had been collected and organized since the 1970s, they donated everything to the LSU Libraries. Many of the materials were federal government documents that the Libraries did not have in paper. The project librarian and the Libraries' government documents librarian acquired approximately 400 grey items from Sea Grant. These items were moved to the Libraries for processing. In addition to the Sea Grant acquisition, two retired LSU professors donated their personal collections to the Libraries. During discovery visits the project staff did not remove any documents from state agencies. All data collected at the agencies was compiled in Excel spreadsheets and sorted by title. The spreadsheets were collocated and the locations where a document could be found were noted in the item record.

Project Questionnaire

In order to find out more information about the research habits and information needs of potential users of the coastal grey literature database, the project librarian distributed a research preferences questionnaire (Appendix A). Questions relating to how the respondents conduct research, which online databases they search, and what types of grey literature they find most helpful comprised the questionnaire.

A description of the project along with a link to the questionnaire was sent via e-mail to 370 individuals who work in coastal-related jobs in Louisiana. Individuals were asked to complete the questionnaire located on the project website. When compiling the list of participants, care was taken to identify individuals working in government agencies, both federal and state; university departments; non-profit agencies; and private companies. The project staff made every effort to contact a proportionate number of people from each group, but discovered that it was more difficult to find contact information for state and federal agency workers than for university faculty and graduate students.

One week after the first e-mail solicitation, nearly 60 individuals had responded to the questionnaire. Two weeks after the first e-mail, a reminder e-mail was sent to only those who had not yet responded. Over 100 individuals, or 34% of the total number solicited, responded. Fifty-one individuals identified themselves as state or federal government agency employees, 50 as university researchers, and four indicated that they work for private companies, nonprofits, or other.

This exercise was not meant to be scientific, but to allow the projected target audience the opportunity to share their research preferences and shape the project outcome. The responses were positive and many contained detailed descriptions of how the individual conducts his/her research. Several had suggestions for other types of grey literature that would be useful for inclusion in the database. Respondent feedback was evaluated and used to better identify grey literature sources and to create the custom database. The project graduate assistant entered results into a Microsoft Access database where the results were queried and tabulated.

Respondents were asked to rank resources they use for their research in the order in which they use them. "Journal articles" was chosen as the number one source by 51 individuals. Approximately seventy-five percent of those 51 respondents were university researchers. "Technical/Government reports" is the primary source for 31 individuals, of which approximately seventy-four percent were state or federal government employees.

The number of people who research using the Internet and books was significantly smaller; only two individuals chose books as their primary resource for research, and twelve chose the Internet as the first place they conduct research. This trend illustrates that researchers are apt to use the most convenient resource available to them first, which seems to be journals for those in the university setting and government reports for those working in government positions.

Many respondents replied in detail to the question "How do you conduct your research? Please explain in as much detail as you would like how you collect information for a project." Answers ranged from beginning a project by using an Internet search engine to conducting a full literature review of the topic. Searching databases that are available through a library website was mentioned more frequently by individuals with access to university library resources. In fact, in later questions, some respondents who were not a part of a university community indicated that they would like to gain access to the databases that are usually only available to faculty or students. A number of individuals indicated that they seek out information informally from their colleagues as the first step in their research process.

When the respondents were asked if there is a limit on the age of the material they use for research, an overwhelming number, 75%, indicated that they would use materials older than 25 years for research. A similar question, "Would more availability or easier access to older research materials encourage you to use them more frequently?" elicited 92 YES responses. Overall the respondents indicated that older material is still relevant to them in their research.

Respondents were asked to mark all grey literature resources that they find useful or valuable from a list or 11 types. Many individuals checked several boxes, indicating that they find a wide variety of grey literature useful. Ninety respondents in all stated that "U.S. government reports" were useful to them; sixty-nine listed "theses and dissertations" as valuable resources; eighty-seven found "state government reports" helpful; and 72 indicated they use "conference proceedings." All respondents checked the box in the last question indicating that they would like to be contacted when the database was ready.

Database and Metadata

The project database was created by programmers working at the LSU Center for Computation and Technology (CCT) using a portal framework software called GridSphere. The relational metadata database uses a Java-based, object-oriented framework. The database design was constructed to be adaptable to the changing technological needs of the project. The data is searchable by title, author, agency, and keyword. Additionally, the keywords were cross-referenced and like terms were linked to each other. For instance clicking on the keyword "dredging**"** would allow a user to find other documents with that keyword.

The data in the database was organized using the Library of Congress's Metadata Object Description Schema (MODS). This XML based markup language is derived from MARC 21, so it works well with existing library records. It also allows for the attachment of electronic files to the records.

Metadata about each document was recorded onsite in Microsoft Excel. Project staff then identified all items that were already in the LSU catalog by adding a "035" MARC field with the phrase "coastaldb" that would not be visible to patrons. The MARC records for the items in the LSU Libraries were extracted from the catalog and then converted to MODS using MARCedit software.

Project staff also searched for identified documents in the state-wide union catalog, which contains records for 30 Louisiana libraries. When looking for the records of items that were found at the state agencies, project staff also attempted to identify all relevant documents in the catalogs. Importing the metadata for approximately 770 records from the LSU Libraries catalog saved the project staff time and allowed them to input data for agency items that were not cataloged.

A surprising but positive finding was that the majority of the documents that the project staff found at the state agencies were already cataloged and available at LSU Libraries and other libraries in the state. About one quarter of the identified documents was only available at the state agency. This indicated that Louisiana librarians had been diligent in collecting and providing access to coastal Louisiana grey literature.

Lessons Learned

This unique project was ambitious and untried, thus there were many missteps along the way. When attempting a project on this scale it is very important to have a thorough and detailed plan, including timelines, individual responsibilities, and measurable goals. Very little pre-planning was done, which meant that the project staff often improvised to find solutions to problems as they arose.

The most difficult obstacle to the success of this project was the actual database construction. The Libraries does not employ programmers on staff, so they sought outside help from LSU CCT. Working with the CCT programmers was often difficult because of scheduling conflicts and too few of their hours dedicated to the database. Complicating the situation was the decision by the CCT supervisor who managed the programmers to change the database programming from MySQL to Java almost four months into production of the database. The project staff belatedly wished that they had explored more "out of the box" options instead of relying on programmers to build a database concurrent to the actual data gathering.

In addition, the project staff discovered the programmers who were fluent in many programming languages did not speak "library language," and therefore did not have prior knowledge of metadata schemas such as MODS or MARC. The two languages often use similar terms with different meanings, which resulted in confusion and several discussions of the same issues. Both the programmers and the project staff were working on their part of the project at the same time, which proved to be very ineffective since the project librarian needed the database well before it was completed.

Future Plans

This project was proposed as a pilot endeavour when GOCA and the Libraries first began negotiating the funding. The Libraries had hoped that this initial phase of the project would yield the citation-only database, and the second phase, which would require additional funding, would include digitizing the grey literature and providing full-text searchable PDF files to users. These plans were drastically altered, though, when Hurricanes Katrina and Rita struck Louisiana in August and September 2005. Priorities in the state shifted quickly and this project was, understandably, not a high priority for GOCA. In addition, most of the contacts that the project staff had been working with were now engrossed in the immediate problems of assessing the damage to coastal Louisiana and proposing how a tragedy of this magnitude could be prevented in the future.

As a result, the grey literature database was not introduced to the public as planned. The project ended while the project staff was still in negotiations with a state-wide consortium called the Coastal Louisiana Ecosystem Assessment and Restoration (CLEAR) to migrate the database to their server. Over a year after Hurricane Katrina, the database is still not available to users because certain elements were never completed by the programmers. Further customization of the database interface, thorough testing of its functionality, and the addition of other identified grey items have yet to be completed. Although the project is no longer funded, LSU Libraries is committed to finding a suitable host for the database. Early in the project, GOCA and the Libraries decided that a state-wide, inter-agency consortium like CLEAR would be the best type of host for the database, because neither the state agencies nor the libraries would maintain control of the information. When the database is in production the project librarian, who now works for the Libraries in another capacity, will re-visit the state agencies, conduct demonstrations, and train users how to use the database.

Conclusion

Although natural disasters and poor initial planning have delayed the completion and launch of the coastal grey literature database, the project is still ongoing. Preserving Louisiana coastal grey literature is an important part of the re-building and maintaining of the coastal region. The need to preserve and protect Louisiana's coastal wetland becomes more urgent with each passing year. When completed, the Louisiana Coastal Grey Literature Project will provide information and scientific resources to all parties interested in the history and future of Louisiana's coastal wetlands.

Appendix A:

QUESTIONNAIRE ABOUT RESEARCH PREFERENCES

Your input is very valuable to the success of the Coastal Research Gray Literature Project. The database of gray literature will be designed for researchers, scientists, and others working to save Louisiana's coastal wetlands. Please take a few minutes and fill out this questionnaire about your research preferences and your knowledge of coastal gray literature. Your responses will help us further tailor the database to meet your needs.

First and last name:

Agency, University or Company name:

E-mail address:

1. What are your primary sources for information in your research/on the job? Please number these items in the order in which you would use them for research.

Journal articles

Technical or Government Reports

Books

Internet resources

Other, Please specify:

2. How do you conduct your research? Please explain in as much detail as you would like how you collect information for a project. Where do you look for/find the material? What is the process?

3. An online database is a reference tool usually found on a library website that contains citations or full-text articles from journals or other resources. If you use online databases, how do you conduct a search? (Choose any that apply.)

I use keywords or subjects

I browse titles and/or subjects

I search by author name

I don't use databases/am not familiar enough with them to say.

4. Please name a particular database or databases that you search frequently.

Why do you prefer this database(s)?

5. Is there a limit on the age of the material you use for research?

No, I will use whatever material is useful

Yes, I prefer to use material that has been published in the last 25 years

Yes, I prefer to use material from the last 5 years

6. Would more availability or easier access to older research materials encourage you to use them more frequently?

Yes

No

7. What types of gray literature do you find useful or valuable? (Check all that apply.)

- ☐ Annual reports
- ☐ Bulletins
- ☐ Committee Reports
- ☐ Conference Proceedings
- ☐ Data Sets
- ☐ Newsletters
- ☐ Working Papers
- ☐ Theses and Dissertations
- ☐ State Plans
- ☐ State Government Report
- ☐ U.S. Government Report

8. Are there additional types of gray literature materials (not mentioned in Question 7) that you would like to find in the Gray Literature database?

9. Please include any comments you have about the gray literature availability and the database project.

10. Would you like to receive an e-mail when the Coastal Gray Literature Database is ready for use?

- ◉ Yes
- ○ No

Thank you for taking the time to give us your valuable feedback.
Please click the Submit button to send your responses to the Project Librarian.

Accessed 11/13/06: http://www.lib.lsu.edu/special/coastal/questionnaire.htm

References

[1] Louisiana Department of Natural Resources (2006), "Louisiana coastal facts," http://dnr.louisiana.gov/crm/coastalfacts.asp.
[2] U.S. Army Corps of Engineers (2004), "The Mississippi River and tributaries project," http://www.mvn.usace.army.mil/pao/bro/misstrib.htm
[3] Ibid.
[4] Louisiana Department of Natural Resources (2006), "Louisiana coastal facts," http://dnr.louisiana.gov/crm/coastalfacts.asp.
[5] America's Wetland, (2006) "Causes of loss," http://www.americaswetlandresources.com/background_facts/detailedstory/causes.html
[6] American's Wetland, (2006) "Hurricanes Katrina and Rita's impact on America's wetland," http://www.americaswetlandresources.com/background_facts/basicfacts/hurricane.html

https://doi.org/10.26069/greynet-2026-000.608-tgj

Harnessing NASA Goddard's Grey Literature: The Power of a Repository Framework*

Nikkia Anderson, Gail Hodge, and Andrea Japzon *(United States)*

Abstract
The NASA Goddard Library collaborated with several projects on-Center to create a framework for the development of web-accessible repositories of grey literature. Tools and methods for collaboration were developed through a series of prototypes with a variety of Goddard projects based on the Library's Digital Asset System (DAS), a repository to describe and provide access to project information including images, videos, web sites, and technical reports.

Metadata is a key component of this framework. The Goddard Core Metadata Element Set, an extension of the Dublin Core, is used to describe these resources. Additional elements include project name, project code and instrument name. A taxonomy of controlled subject terms has been developed which can also be extended and tailored for each project. This provides consistent searching across the DAS, while at the same time enhancing the search precision for each project collection when it is accessed as a separate collection in the DAS or through its own web site.

To-date, the Library has used the framework with a number of different projects. A commemorative CD with key documents, web sites and oral histories by the project managers was developed for the Hitchhiker Project as it was being disbanded. Metadata from the Swift project library was transformed and made more accessible with pointers to documents in the Swift project library. The Landsat Legacy Project, a joint project with the Landsat Program Science Office, the US Geological Survey, and the NASA History Office, is creating an archive of essential technical-, policy-, and science-related documentation. To gather significant documentation from the over 35-year-multi-agency history of Landsat, the Library added components for external submission of documents for potential inclusion in the new repository and for scanning paper documents. In addition, video histories are being captured from Landsat veterans.

Through these projects, the NASA Goddard Library has developed a methodology for collaborating with different kinds of projects. It involves procedures for analyzing the needs of a project and determining how the metadata, taxonomy, and interface might need to change, while remaining consistent with the DAS framework for cross-repository access.

Ultimately, the Library plans to extend the effort to other projects. (At any given point in time, Goddard has more than 30 projects in various stages of completion.) As part of Goddard's knowledge management activities, the DAS provides a framework for sharing grey literature that would otherwise be scattered across independent project libraries. Benefits include the ability to more quickly find and reuse information to decrease project costs, enhance safety and promote innovation.

Developing the Framework:
In 2001, the NASA Goddard Library began its efforts to develop a center-wide metadata repository of technical and scientific information related to the mission of the NASA Goddard Space Flight Center (GSFC). Through those efforts the Library has examined various approaches to making information accessible and available for re-use. An in-house prototype infrastructure for creating a combined metadata repository to describe and provide access to project information including images, videos, web sites, and technical reports was developed and called the Digital Asset System (DAS). The DAS was developed against the Goddard Core Metadata Element Set, based on qualified Dublin Core [1] with extensions. The Goddard Core Metadata Set was specifically developed to provide better discovery and evaluation in the

* First published in the GL8 Conference Proceedings, February 2007

Goddard context of project management [2], i.e., helping the user find the document, evaluate its usefulness and locate it whether in paper or in digital form. The Goddard Core was developed through the work of the Goddard Metadata Review Group (MRG), which consist of internal and external metadata experts with the primary focus on metadata for project documentation and related objects. The MRG supports the development of a metadata framework for digital archiving, preservation, and access across objects and subjects to advance knowledge management efforts [3].

To expand the elements of findability, the Goddard Library incorporated the NASA Taxonomy and EOS (Earth Observatory System) Taxonomy as controlled discipline terminologies for the subject element of the Goddard Core. The NASA Taxonomy was developed through a NASA Headquarters contract to the Jet Propulsion Laboratory (JPL), another NASA center located in California, in an effort to provide a strategic search mechanism across the NASA centers, thereby building a knowledge base to assist in browsing and navigation through large collections of disparate information objects [4]. The EOS Taxonomy was developed at NASA GSFC to help support and serve the EOS Program Office for classifying information on a Web-portal [5]. In our efforts to increase the search precision over the years, while working with specific projects to gain access to important documents which no one could find, we developed project specific controlled vocabularies to bring those once grey documents to life.

Projects and Project Libraries (Home of the Grey Literature)

NASA GSFC has at least 40 active missions in various life cycle phases, at least 10 missions accomplished, and at least 15 missions planned to orbit. During these missions, NASA GSFC produces a tremendous amount of information. Some of these missions have project libraries that keep track of the information that pertains to the unique engineering and project management tasks involved in designing, implementing, launching, and maintaining the spacecraft [5]. There are some missions that don't have project libraries and the information is scattered about in various hands of the scientists, engineers, and project managers involved in the missions with little or no likelihood of being available for further use.

Through the Goddard Library's knowledge management initiatives, the Library became the face of resource discovery and information re-use in the project science and engineering arena at Goddard.

The Library was already known for providing assistance in finding information to suit the needs of its users, which are primarily the 9,000 employees (civil servants and contractors) on Center. Now with the knowledge management role, the Library was able to use other resources such as the Goddard Core Metadata Element Set and the DAS to help the users search information consistently across various project resources and not have to familiarize themselves with each individual project search system when available.

Starting Simple: The Hitchhiker CD

The venture into bringing grey information to life began with the Hitchhiker Project. The Hitchhiker Project began in 1984 and launched in 1986 [6]. In 2001, as the Hitchhiker mission was being hibernated, the Chief Knowledge Management Officer asked the Library to compile the information of the Small Shuttle Payloads Project to share with the project team and others interested in the project. Unlike other project collaborations that came after Hitchhiker, the Library didn't get a chance to create a sub-system of the DAS for Hitchhiker; instead we created a commemorative CD with web sites, images, key project documentation, and oral histories to serve as a finding aid for key information pertaining to the mission (see Figure 1).

Hitchhiker helped us see what was needed to develop sub-systems of the repository. Outlining the project scope was definitely a big need. When working on Hitchhiker, the Library was often at a loss, because the project team was no longer available to give the background of the mission and lessons learned. The Library was able to interview a few of the scientists and capture those interviews which were documented and made accessible on the resulting CD.

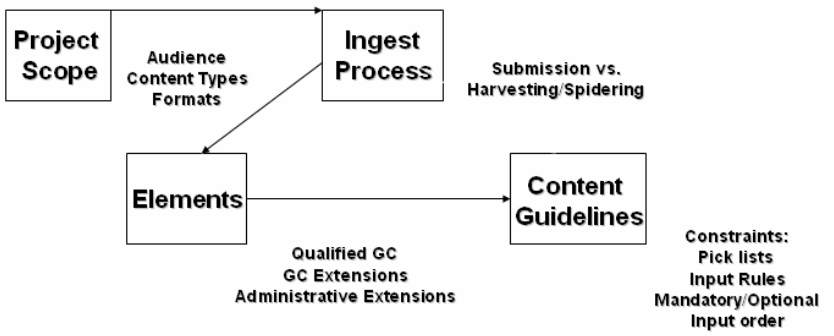


Figure 1: DAS Framework

Harvesting and Mapping a Project Library's Metadata: The SWIFT Metadata Repository

The next effort at bringing the grey literature to light involved the Swift Metadata Repository and modifying the infrastructure of various project resources such as Goddard Project Directory, Goddard Project Libraries Database, and the IMAGES collection developed by the Goddard Library outside the DAS. During this time the framework went from a "what if" to a formalized way of searching and providing access to grey literature. The Swift mission launched in 2002 and is still in orbit with a focus on gamma ray burst. The Swift Metadata Repository was developed as a sub-system of the DAS with images of the Swift launch, web sites, videos, and pointers to project documents. The Swift project library uses a Centralized Configuration Management System (CCMS) for submission of documents, but the drawback is that the CCMS does not have a user-friendly search interface. In order to get information that you are looking for you must know as much as you can about that document to guarantee findability. So the Library took on the challenge of providing pointers with key information that would improve the access to the documents from the CCMS.

Swift Metadata Elements	GC Metadata Elements
Number	Identifier.Original
Payload	Subject.MissionProject/Subject.Instrument
Responsible Person/Organization	Creator.Employee/Creator.Organization
Title	Title
Alt Number	
WBS Number	
Sub System	Subject.Instrument
Revision/CH/DCN	
Submitted Date	
Approved Date	Date.Available
Approver	
Remarks	

Figure 2: Swift Metadata Mapping to the Goddard Core

The Library mapped the metadata of the CCMS to the Goddard Core Metadata Element Set (see Figure 2) and incorporated other metadata elements required for the new content types. The Library developed a controlled vocabulary of terms specific to the Swift mission and customized some of the elements to project-specific controlled vocabularies. The user can simultaneously search across various content types using a single search box. Searching the metadata stored in the DAS retrieves the document which can be inserted into the CCMS search box to retrieve the document of choice. This system helps users discover the information resources of the Swift project through a single interface rather than separately searching for images, videos and documents.

Adding Ingest Options and Other Content Types: The Landsat Legacy Project
In 2004, the Goddard Library was contacted by the Landsat Program Science Office to help them create a Landsat document archive. The type of documentation sought for the Landsat archive includes policy-, technical-, and science-related materials with an emphasis on internal technical papers (or grey-literature) of both the government and the aerospace industry. The project has been dubbed the Landsat Legacy [7]. The Landsat Legacy team consists of a joint collaboration with the Landsat Program Science Office, the US Geological Survey, and the NASA Goddard Library [8]. The Landsat Program has been in existence since 1972 when Landsat 1 was launched. Currently the LPSO is on its 7th Landsat mission (Landsat 7). The DAS framework was incorporated to prototype this sub-system just as the others described above but with a few enhancements.

The original scope of this project was to include the policy-, technical-, and science-related documents but with further discussion regarding the possible deepness of the collection the idea struck of having video histories. The Library took on the role of recording live interviews of veterans of the Landsat Program over the last 35 years. By including the video histories to the scope of the project, the Library still continued to follow the process of the framework, which included selecting appropriate elements from the Goddard Core for the infrastructure and developing what is now the largest controlled vocabulary that we've created to date. The controlled vocabulary took on many versions before we got close to finalizing but with project-specific controlled vocabulary there is always room to expand. Drop down menus of controlled vocabulary for instrument, audience, and mission project were also added.

To gather significant documentation from over 35 years of the multi-agency history of Landsat, the ingest process of the framework was expanded to include additional components. The Library added components for external submission of documents for potential inclusion in the new repository and for scanning paper documents. The external submission of documents was set up to allow donors to submit documents through a web-interface that they may have stowed away in old filing cabinets or in boxes in there garage or home offices. Although the system is web-based the submission could be either electronic or paper documents, which allowed the Library to advance the DAS to include scanning of project documentation. The scanning process required procedures to be developed to ensure consistency and quality regardless of who did the scanning.

Since 2004 the Library has continued to make further advancements in the creation of the sub-system repository for the Landsat Legacy Project. Like the other DAS projects, the Legacy System will also simultaneously search across web sites, documents, and video histories. There may be efforts in the future to expand the project scope to provide access to the Landsat data archive for scientific images.

Conclusion – The Power of the Repository
Through these projects, the NASA Goddard Library has developed a methodology for collaborating with different kinds of projects. It involves procedures for analyzing the needs of a project and determining how the metadata, taxonomy, and interface might need to change, while remaining consistent with the DAS framework for cross-repository access. The framework allows for expanding in all areas as described through each project prototype. Through a series of challenging projects, the Library has improved the DAS and expanded the Goddard Core. The DAS and the Goddard Core have been presented as accomplishments through special presentations to the NASA Web Managers Group and the Metadata Review Group [9]. Through its efforts to harness the grey literature of NASA Goddard through a repository, the Library continues to provide consistent access, to share information across projects and to advance Goddard's knowledge management initiatives.

Acknowledgements
This work was conducted under NASA Goddard Contract NAS5-01161. The authors gratefully acknowledge the support and encouragement of Robin Dixon of the NASA Goddard Space Flight Center Library.

Works Cited

[1] Dublin Core Metadata Initiative. "Dublin Core Metadata Element Set 1.1." [Online]. Available: <http://www.dublincore.org/documents/dces/> [November 6, 2006]

[2] Senserini, A., Allen, R., et al "Archiving and Accessing Web Pages: The Goddard Library Web Capture Project." D-Lib Magazine. [Online] Available <http://www.dlib.org/dlib/november04/hodge/11hodge.html> [November 2004].

[3] Metadata Review Group [Online]. Available: <http://library.gsfc.nasa.gov/mrg> [March 3, 2006]

[4] Busch, J. (2004). "Taxonomy Use Examples." [Online] Available: <http://library.gsfc.nasa.gov/mrg/pdf/NASA_Taxonomy.pdf> [October 28, 2004]

[5] Hodge, G., T. Templeton, et al "A Metadata Element Set for Project Documentation." Science & Technology Libraries, Vol. 25(4) [2005]

[6] NASA Facts Sheet. "The Hitchhiker Project." [Online]. Available: <http://www.gsfc.nasa.gov/gsfc/service/gallery/fact_sheets/general/hitchh.htm> [November 13, 2006]

[7] Rocchio, L., G. Hodge, T. Arvidson, D. Williams, and J. Irons. The Landsat Legacy: Tracking down three decades of knowledge. In: Pecora 16 "Global Priorities in Land Remote Sensing" Proceedings. American Society for Photogrammetry and Remote Sensing, Sioux Falls, SD. [2005]

[8] Landsat Legacy Document Registry. [Online]. Available: <http://library.gsfc.nasa.gov/landsat> [July 25, 2005]

[9] Anderson, N. & G. Hodge. "The Goddard Core: A Metadata Framework for Web Site and Beyond - Presentation at Web ManagerViTS." [Online]. Available: <http://library.gsfc.nasa.gov/mrg/pdf/GC_WebManagersViTS_07242006.pdf> [July 24, 2006]

https://doi.org/10.26069/greynet-2026-000.608-tgj

Hyperactive Grey Objects*

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

Abstract
Previous papers on Grey literature by the authors have described (1) the need for formal metadata to allow machine understanding and therefore scalable operations; (2) the enhancement of repositories of grey (and other) e-publications by linking with CRIS (Current Research Information Systems); (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. This paper takes the development one step further and proposes 'intelligent' grey objects.

The hypothesis is in 2 parts: (1) that the use of passive catalogs of metadata does not scale (a) in a highly distributed environment with millions of nodes and (b) with vastly increased volumes of R&D output grey publications with associated metadata; (2) that a new paradigm is required that (a) integrates grey with white literature and other R&D outputs such as software, data, products and patents (b) in a self-managing, self-optimising way and that this paradigm manages automatically curation, provenance digital rights, trust, security and privacy.

Concerning (1) existing repositories provide catalogs; harvesting takes increasing time ensuring non-currency. The end-user expends much manual effort / intelligence to utilise the results. The elapsed time of (i) the network (ii) the centralised (or centrally controlled distributed) catalog server searches (iii) end-user intervention becomes unacceptable.

Concerning (2) there is no paradigm currently known to the authors that satisfies the requirement. Our proposal is outlined below.

Hyperactive combines both hyperlinking and active properties of a (grey) object. Hyperlinking implies multimedia components linked to form the object and also external links to other resources. The term active implies that objects do not lie passively in a repository to be retrieved by end–users. They 'get a life' and the object moves through the network knowing where it is going.

A hyperactive grey object is wrapped by its (incrementally recorded) formal metadata and an associated (software) agent. It moves through process steps such as initial concept, authoring, reviewing and depositing in a repository. The workflow is based on the rules and information in the corporate data repository with which the agent interacts. Once the object is deposited, the agent associated with it actively pushes the object to the end-users (or systems) whose metadata indicate interest or an obligation in a workflowed process. The agents check the object and user (or system) metadata for rights, privacy, security parameters and for any charges and assure compatibility.

Alternatively the object can be found passively by end-user or system agents.

The object can also associate itself with other objects forming relationships utilising metadata or content. Declared relationships include references and citations; workflowed relationships include versions and also links to corporate information and research datasets and software; inferenced relationships are discovered relationships such as between documents by different authors developed from an earlier idea of a third author.

Components of this paradigm have been implemented to some extent. The challenge is implementing – respecting part two of the hypothesis - the integration architecture.

* First published in the GL8 Conference Proceedings, February 2007

Background

Previous papers on Grey literature by the authors (in the GL Conference Series) have described:

a) the need for formal metadata to allow machine understanding and therefore scalable operations (Jeffery 1999);
b) the enhancement of repositories of grey (and other) e-publications by linking with CRIS (Current Research Information Systems) (Jeffery and Asserson 2004);
c) 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).

This paper takes the development one step further and proposes 'intelligent' grey objects.

The Hypothesis

The hypothesis is in 2 parts. The first part describes the problem, the second our proposed solution. The problem is that that the use of passive catalogs of machine readable (but not machine understandable) metadata (DC) does not scale, particularly in a highly distributed environment with millions of nodes and with vastly increased volumes of R&D output grey (and white) publications with their associated metadata. The solution is that a new paradigm is required that integrates grey with white (hypermedia) literature and other R&D outputs such as software, data, products and patents, in a self-managing, self-optimising way and that this paradigm manages automatically curation, provenance digital rights, trust, security and privacy.

State of the Art

Concerning the problem outlined in the previous section:

a) Existing repositories catalogs have a high cost of human effort in input / update;
b) Harvesting takes increasing time - ensuring non-currency of the repository;
c) To obtain and utilise results the end-user expends much manual effort / intelligence because the metadata is machine-readable but not machine-understandable;
d) The elapsed time of the network, of the centralised (or centrally controlled but distributed) catalog server searches and of end-user intervention becomes unacceptable.

Concerning the solution, there is no paradigm currently known to the authors that satisfies the requirement - hence our proposal has been developed. However the existence independently of components of the proposed solution provides encouragement and optimism that the solution is feasible.

The Notion

The notion is of a hyperactive object, an encapsulated organism. The object can be a textual document, a multimedia document or a hypermedia document. Its properties combine both hyperlinking and active properties of a (grey) object. Hyperlinking implies multimedia components linked (composed) to form the object and also external links to other resources. The term active implies that objects do not (only) lie passively in a repository to be retrieved by end–users. This concept has been explored in various projects, one example being (AOP) and there are many others e.g. the projects from (Softagents). Taking it one step further, the objects 'get a life' and the object moves through the network knowing where it is going (Codeproject); this owes something to bio-organisms for its inspiration.

The hyperactive notion – combining the active and hyperlinked properties of the object - originates from early work at CCLRC on linking workflow and complex hypermedia research documents evolving in the early nineties to consideration of utilising a WWW environment and followed by later work in the mid-nineties on active retrieval, push technologies and filtering. The context of this work was the intranet office automation and e-business-process environment, not – at that time - the e-library or grey literature environment. In parallel other teams were working along similar lines, for example (Sanchez, Leggett and Schnase 1994). Developing thought since then leads to the current concept proposal represented diagrammatically (Figure 1).

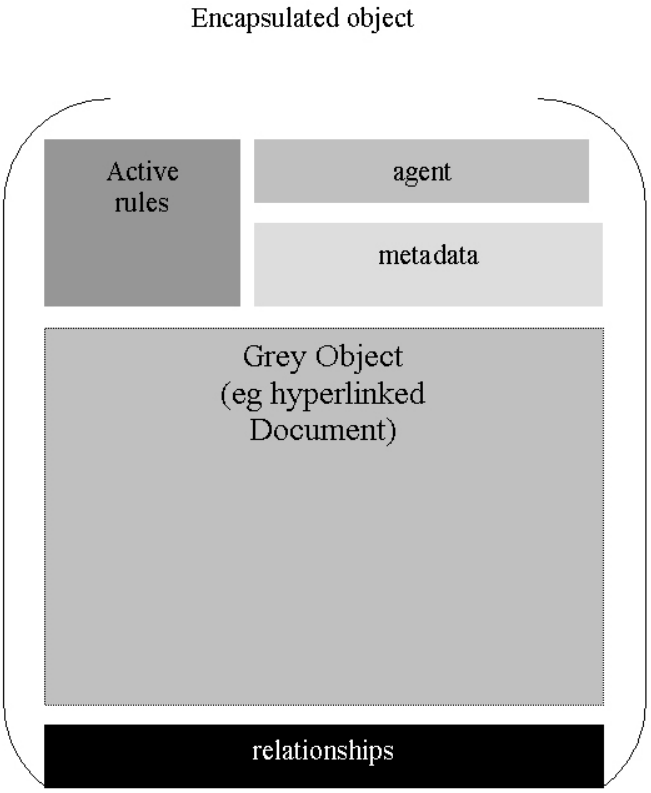


Figure 1

How it works

A hyperactive grey object is wrapped by its - incrementally recorded - formal metadata (schema, descriptive and restrictive as in (Jeffery 2000)), its active rules (Paton 1999), its relationships (e.g. references, citations) and an associated software agent (Wikipedia: Software Agent). It moves through process steps such as initial concept, authoring, reviewing and depositing in a repository. It should be noted that this ensures a low effort threshold for metadata collection (Jeffery and Asserson 2005).

The workflow is based on the rules and information in the CDR (Corporate Data Repository). This is the 'nerve centre' controlling an organization by recording the relationships - including authorities and permissions - between objects. The agent interacts with the CDR and with the active rules within the encapsulated grey object - at least some of which are usually downloaded from the CDR to control the workflow. Once the object is deposited in the repository, the agent associated with it actively checks and builds the relationships (hyperlinks) to make them active. The agent then, after review and at the time of publication, pushes the object to the end-users - or systems - whose metadata indicate interest or an obligation in a workflowed process. The agents check the object and user - or system - restrictive metadata for rights, privacy, security parameters and for any charges and assure compatibility before making the desired object available.

Alternatively, of course, the object can be found passively by end-user or system agents using its descriptive metadata for retrieval / discovery. Again, in this case, the agents check the restrictive metadata. The overall architecture is represented in (Figure 2).

Wider Context

The object can also associate itself with other objects forming relationships utilising metadata or content. There are three main types of such relationships - which can also be considered as hyperlinks if the semantics are discarded:

a) Declared relationships including references and citations;
b) Workflowed relationships including versions and also links to corporate information and research datasets and software;

c) Inferenced relationships which are discovered relationships such as between documents by different authors developed from an earlier idea of a third author.

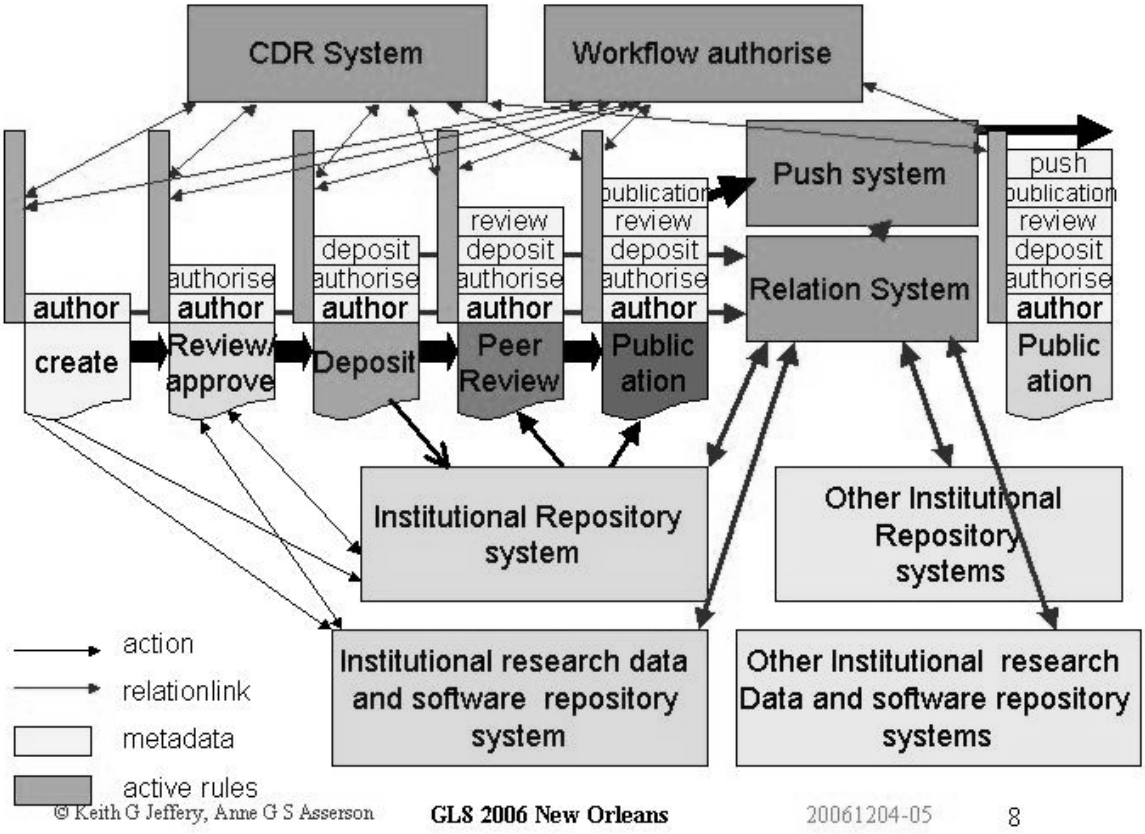


Figure 2

Current State

Components of an architecture supporting this paradigm have been implemented to some extent. Formalised metadata (and links to CRIS: Current Research Information System) has been defined (Asserson and Jerffery 2004) and part-implemented at CCLRC using (CERIF) as the CDR. Metadata linking of a CRIS (FRIDA) to an Institutional Repository (DSpace) is being implemented at UiB. Workflow using the CDR is in production at CCLRC. Hyperlinking of objects by manual procedures has been a production facility for years; automating it is not so easy but using CRIS systems as metadata the formal relationships in the CRIS using (CERIF) make this possible. Inferencing - usually for plagiarism detection in this domain - has been demonstrated in several contexts and is known technology. Agents have been demonstrated in many research prototypes - as have active rules – during and since the eighties. The challenge is implementing – respecting part two of the hypothesis - the integration architecture, including the self-* (self-managing, self-tuning, self-scheduling, self-repairing) properties. This requires GRIDs technology (Jeffery 2004), (Jeffery 2004a).

Conclusion

We have proposed a solution to the scalability problem of current environments which require too much human intervention and are insufficiently amenable to intelligent machine behaviour, largely due to the lack of formal metadata. Our solution has the desirable properties of minimum effort by the end-user and maximum effort by the system with maximum use of the available information by the system to manage workflow, to manage relationships, to generate relationships and to assist retrieval in a scalable way.

References

(AOP) http://www.oberon.ethz.ch/active/

(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)

(CERIF) www.eurocris.org/cerif
(Codeproject) http://www.codeproject.com/gen/design/MobileAgents.asp

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

(DSpace) http://www.dspace.org/

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

(Jeffery 1999) Jeffery, K G: 'An Architecture for Grey Literature in a R&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 2004) 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 2004a) 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 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

(Paton 1999) Norman Paton (Ed) Active Rules in Database Systems. Springer, New York 1999, ISBN 0-387-98529-8

(Sanchez, Leggett and Schnase 1994) HyperActive: extending an open hypermedia architecture to support agency ACM Transactions on Computer-Human Interaction (TOCHI) Volume 1 , Issue 4 (December 1994) Pages: 357 - 382 1994 ISSN:1073-0516

(Softagents) http://www.cs.cmu.edu/~softagents/

(Wikipedia: Software Agent) http://en.wikipedia.org/wiki/Software_agent

https://doi.org/10.26069/greynet-2026-000.608-tgj

A Public Health Knowledge Management Repository that Includes Grey Literature*

Debra Revere, Paul F. Bugni, and Sherrilynne Fuller (*United States)*

Abstract
Problem: Public health professionals rely heavily on resources that are often only available in grey literature format. However, while grey literature may contain comprehensive, concrete, and up-to-date information, the fugitive nature of this material makes access problematic. The public health community needs a knowledge management repository of grey literature and tools for easy and rapid access, so time spent searching across and through materials can be reduced.
Goal: Design a customizable prototype public health knowledge management repository system and end-user interface with optimal interoperability and the capability to provide timely access to public health information in support of decision making at the point and time of need. Specification of an appropriate metadata schema, which identifies in a standardized way the elements needed to describe a resource, are a critical part of the system. The long-term goal is a system that delivers answers to public health questions, not a list of pointers to resources that may or may not contain information to answer those questions
Evaluation Procedure: We are utilizing user needs analysis, user profiling, and resource assessment to inform understanding the information needs of public health professionals in the context of their everyday workflow and enable identification of key grey literature knowledge resources for incorporation into the knowledge management system. Rapid prototyping is being used to translate these findings into system specifications and interface design of a small-scale prototype system. The prototype defines system components and interactivity both among components and with relevant external knowledge resources—for example, the New York Academy of Medicine's Grey Literature collection, web resources from the Centers for Disease Control and Prevention, Department of Health materials, etc. The collection of materials will be organized utilizing resource metadata (high level formal, standards-based descriptions of documents) to improve location of relevant grey literature and other information sources.
Results: Testing and evaluation will result in enhancements to the user interface, information resources, presentation of those resources, etc. We also anticipate that the metadata schema employed in a public health knowledge management system will improve the efficacy and efficiency of locating answers to public health questions from the grey literature.
Conclusions: As the amount and breadth of public health information resources continue to expand it is critical that we find ways to provide direct access to the contents of these rich and complex resources. We believe that a public health grey literature knowledge management system with a collection of resources driven by the information needs of public health practitioners and organized using an appropriate metadata scheme will reduce time spent searching across and through materials, enhance public health decision making and ultimately improve the overall quality of public health services.

Introduction

Public health is a broad interdisciplinary field that relies on a variety of information sources that are largely uncontrolled. Accurate and timely information access and distribution is central to effective and proactive public health practice. Data, reports, studies, guidelines, etc. produced by private organizations and public agencies at local, state, national, and international levels are commonly not managed in any systematic or comprehensive way at the present time. Locating and retrieving these materials at the point and time of need is problematic for public health practitioners.

In October 2005, the University of Washington was awarded one of the first Centers for Disease Control and Prevention (CDC) grants to establish a Center of Excellence in

* First published in the GL8 Conference Proceedings, February 2007

Public Health Informatics (CEPHI) with the mission to improve the public's health through discovery, innovation, and research related to health information and information technology. The myPublicHealth Project is one of two research projects funded by CEPHI. The goal of myPublicHealth (myPH) is the design and development of an interactive digital knowledge management system to support the collection, management, and retrieval of public health documents, data, learning objects, and tools. Public health grey literature is a major information source of myPH's knowledge management repository.

myPublicHealth Goals

The long-term goal of myPH is the implementation of a successful knowledge management system that is tailored to the public health practitioner's information needs, work processes, and environment. The knowledge management repository system aims to improve access to and use of digital information resources in support of evidence-based public health practice.

The current initial design focus is on a prototype public health knowledge management repository system and end-user interface with optimal interoperability and the capability to provide timely access to public health information in support of decision making at the point and time of need. Specification of an appropriate metadata schema, which identifies in a standardized way the elements needed to describe a resource, is a critical part of the system.

Methods

Achieving these goals requires a comprehensive understanding of the information needs, information seeking behavior, and human-computer interaction of public health practitioners. Prior to the system design stage, the myPH implementation plan included: (1) a comprehensive literature review focused on understanding the unique information needs of myPH users [1]; (2) an inventory of public health information resources in use by practitioner, including grey literature; (3) survey/analysis of existing content management systems and tools; (4) system and user requirements; and (5) a workflow analysis in a variety of public health settings. At the time of writing this paper, the first three tasks are completed; the final two tasks are in process.

Literature Review. The systematic literature review, currently in press [1], focused on the information needs of public health professionals prior to developing system requirements to inform the design and development process of myPublicHealth. From this review we learned that selection of public health information resources is influenced by job function, disciplines and training; public health workers need accessible, free or low cost, and stable resources; workforce roles and functions often overlap; and people (e.g., colleagues, personnel) are considered the most reliable resources.

These findings are guiding prototype design in ensuring that the public health knowledge management repository system and end-user interface needs to: reflect the complexity and diversity of the public health workforce itself while providing a user-friendly interface; reduce barriers to access with a minimum of security requirements; and provide timely access to human communication networks (e.g., providing accurate directories, listservs, etc.).

Information Resources Inventory. We collated resources from a variety of sources, including public health practitioner interviews and suggestions, information resources and sources as identified in the literature review, and preliminary findings of the workflow assessment. We are using this inventory to guide requirements for the repository metadata schema.

Survey of Content Management Systems. We reviewed the features of current content management and access systems, including local systems in use at the University of Washington and open-source systems (including Plone, Alfresco, and WIKIs). Evaluation of these systems included their capabilities in the following: ability to integrate with other systems; ease of use; documentation; customizability; user interface; costs; ability to handle multiple document formats; and security issues. At the time of publication Plone has been chosen as the most promising content management system and is being evaluated for integration with locally developed system modules.

Systems and User Requirements. Rapid prototyping and iterative design is a model that incorporates the user in the design of the system. We have already begun using this approach in a small-scale prototype system in order to test out key features of the design and make changes as needed.

At the time of publication, planning is still in progress, but certain tasks and deliverables have been identified, including: a standards-based knowledge repository system (full-text documents or links to full-text documents when available as e-resources); integrative web services to support and enhance retrieval of critical information in support of decision-making by public health professionals; an interface design that allows access to the diverse set of information resources and tools to support reportable disease and other types of public health investigations.

Workflow Analysis. Understanding how the tools can be incorporated into the work environment of the users is an important task. The literature on implementation of computerized clinical systems has consistently reported that neglecting to take user values, needs, and practices into consideration can lead to failed systems, lack of adoption by users, and unanticipated costs. [2-4]

Incorporating Grey Literature into the Knowledge Management Repository System

myPH system architecture will consist of three tiers: content repositories (standard and static); middleware for gathering metadata and wrappers integrating the back-end components; and the front-end user interface. It is anticipated that the static repository will contain the majority of the grey literature. The user interface will include basic and advanced search capabilities, including the ability to search by metadata (for example, title, author, abstract, document summaries, etc.).

Specification of an appropriate metadata schema, which identifies the elements needed to describe a resource in a standardized way, is being used in the design, building and populating of the knowledge management repository. We are incorporating metadata elements from existing metadata standards, for example Dublin Core [5], the UK National Public Health Language Thesaurus [6], and NLM's Unified Medical Language System [7]. These standards will be the schema through which commonalities and mappings are established for sharing information about existing, distributed resources. In addition, we will adhere to the CDC's Public Health Information Network Messaging System (PHINMS) vocabulary standards and specifications. [8]

The standard repository will contain the templates and server-side modifications for processing user requests to access content such as the New York Academy of Medicine's Grey Literature collection and web resources from the Centers for Disease Control and Prevention. The static repository will contain harvested grey literature in the form of XML files with metadata records and repository information.

myPublicHealth Prototype and Preliminary Evaluation

At present, a myPH prototype (Fig. 1) built in flat HTML is hosted online, primarily to gather user feedback on design and content selection.

Tabs across the top of the interface (Fig. 1A) will provide access to resource "toolkits" organized by public health content areas. Resources will also be organized by roles (Fig. 1B). Specialized searching services (Fig 1C) are accessible through the numerous search boxes for querying directly into resources.

As mentioned previously, we will be utilizing user needs analysis, user profiling, and resource assessment to better provide services that meet the information needs of public health professionals in the context of their everyday work. Testing and evaluation will likely result in modifications to the user interface, information resources, presentation of those resources, etc. We will be testing the ability of the system to improve the efficacy and efficiency of locating public health resources.

https://doi.org/10.26069/greynet-2026-000.608-tgj

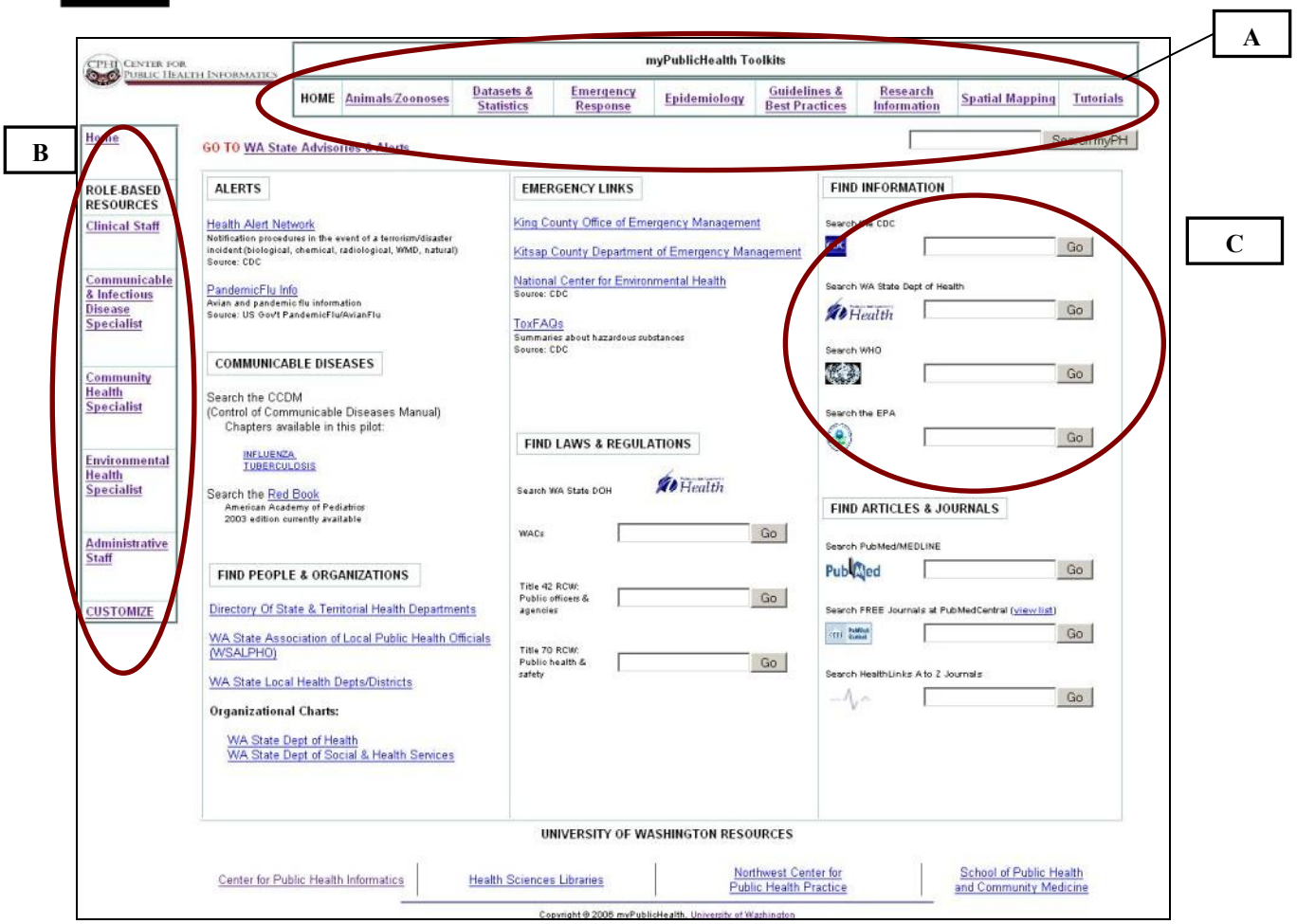


Figure 1. myPublicHealth Prototype, v.2.3

Conclusions

We believe that a state of the art public health knowledge management system that is integrated into a larger public health response system will improve the efficacy and efficiency of locating and applying relevant information to decision making by public health researchers and professionals. Creating such a knowledge management repository and related tools, integrated into other resources that are routinely accessed by public health practitioners, could reduce time spent searching across and through materials, thereby enhancing public health decision making and ultimately improving the overall quality of public health services.

Acknowledgements

Many people contributed to the work that is reported in this paper. We wish to acknowledge the contributions of the myPublicHealth Project team, including Yuki Durham, AnnMarie Kimball, John Kobayashi, Ann Madhavan, Neil Rambo, and Anne Turner. Funding for myPublicHealth is provided by the Centers for Disease Control and Prevention Award Number 1 P01 CD000261-01 which supports the Center for Public Health Informatics in the School of Public Health, University of Washington, Seattle, WA, USA. Additional support was provided by the UW Health Sciences Libraries.

References

1. Revere D, Turner A, Madhavan A, Rambo N, Bugni PF, Kimball AM, Fuller SS. Understanding the information needs of public health practitioners: A literature review to inform design of an interactive digital knowledge management system. *J of Biomedical Informatics*. To appear in Special Issue on Public Health Informatics, Spring 2007.
2. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc*. 2004;11*(2)*:104-12.
3. Kirah A, Fuson C, Grudin J, Feldman E. Ethnography for Software Development. In R Bias, D Mayhew, eds. Cost Justifying Usablity: An Update for the Internet Age, 2nd ed. Elsevier; 2005:415-45.
4. Markus ML. Technochange management: using IT to drive organizational change. *J of Inform Tech*. 2004;19*(1)*; 4-20.
5. Dublin Core Metadata Element Set. Dublin Core Metadata Initiative. <http://dublincore.org/documents/dces/>.
6. National Public Health Language Thesaurus. National Health Service, UK. <http://www.nphl.nhs.uk/>.
7. Unified Medical Language System. National Library of Medicine. <http://www.nlm.nih.gov/research/umls/umlsmain.html>.
8. Public Health Information Network Messaging System (PHINMS). CDC. <www.cdc.gov/phin/software-solutions/phinms/index.html>.

https://doi.org/10.26069/greynet-2026-000.608-tgj

Indexing grey resources: considering the usual behaviour of library users and the use of Dublin Core metadata using a database of specialised vocabulary*

Jacques Cuvillier (*France)*

Abstract
Grey resources spread over a large set of matters. In each field of interest, people have their own concepts, each group will use their own vocabulary. Most authors of Grey resources were not able to determine the Decimal or Dewey classification themselves, nor are they able to fix themselves the appropriate terms to embed in the DC elements. It is then worth providing the webuser with an interface that makes him feel at home, by adapting automatically the terms designing the fields of the search form to those which are in use in his domain of activity. The whole classification proposed cover about 110 different topics, each one having a set of terms chosen to entitle the fields of the interface of thematic search as well the submission forms. Moreover, seven entry points are proposed in the first step of the connection, driving the user towards the appropriate context in accordance with the purpose of the document to find : teaching and research ; literature ; technological documentation and data sheets ; vulgarization and everyday life ; citizen and social activities ; arts , sports and spare time ; bank of images. Indeed, the system must allow the flexibility that allow the replacement of any term by another if necessary. So the entire set of terms is put in a database. This system is also able to adapt the language chosen by the user. The submission forms filled by authors are parsed according to the topic selected and the Dublin Core metadata file - as an XML file - is generated and stored in the indexing server. It is delivered on simple metadata request, and is reachable by an hyperlink included in any page of search results. The experimental system has been built and evaluated over a period of five years, and was implemented in the site of the university of Poitiers since June 2005. The software is to be released under free software license (CECILL).

Introduction
Sharing Grey resources leads to conciliate the role of several partners : authors, library keepers and library users. In the following, the three will be considered as webusers. Regarding authors and library keepers, the purpose of collecting a significant amount of Grey literature - even from authors among whom a part have various concerns for rigorous metadata - leads to conceive an adapted system that makes the indexation process as non-constraining as possible. Moreover, we should keep in mind that classifying a document or embedding metadata within DC elements are not tasks that everyone could manage. Nevertheless, filling up the set of basic Dublin Core metadata is considered as a the lowest goal. The method used for collecting these metadata should naturally lead to convergent interpretations. Regarding library users, the search task has different aspects depending on the widespread use of concerned documents. In a narrow scope, one or two fields of metadata should be sufficient, for example the author's name or the reference of a course. Also, the screens used for document search are up to the library keepers, and smart means of building up on-demand interfaces[1] have been implemented on the site <Document Libre>. Nevertheless, thematic search based on the full set of available metadata is considered as the typical and more powerful solution, and will be considered in the following.

Grey resources and Grey behavior
Presenting to library users a convenient way to find the required resources is our main concern. A comparison can be done about the ways they got in use till now whether they are stepping in a real library or using an Internet browser.
When he steps in a real library, the customer refines his selection step by step : finding the "right store", where the targeted item should be, and then the "right

* First published in the GL7 Conference Proceedings, January 2006

bookshelf" which might group the documents of a given scope. Usually, the bookshelf is named with words closely related to the topic, according also to the public for which the documents are intended. All is done for the customer to feel confident.
Searching via an Internet browser, the webuser enters some keywords in a search engine that returns a tremendous number of suggestions and has to filter the results in a rather randomly manner. Indeed, he should make better use of a specialized website allowing a thematic search. We will examine the means of conciliating the user's wellbeing and the constraints of efficient indexation system.

How to conciliate Grey behavior and efficient indexation?

"Grey literature has been studied and discussed since the 1970s but users are rarely mentioned ... a lack of widespread and consistent cataloging negatively impacts grey literature use." [2]

The wellbeing of authors, library keepers and users is only a part of the goal. The essential is the availability of resources and the relevance of the indexing system. The exploration of this problem started in the very beginning of the last decade[3] but the approach must now take advantage of standardized metadata.
We are aware that Dublin Core metadata brought a decisive stride forward in the area of resources sharing[4], and is of peculiar interest regarding Grey resources. As most Grey repositories are too small to be able to satisfy customers of their own, DC metadata allow to spread the research over a federative network provided they stay in a compatible framework and develop along a concerted and coherent process. Nevertheless, the witty use of this system belongs to experts, and its misuse would lead to helplessness.
The solution suggested here is based on know-how and skills of contributors nested in specialized contexts, and on a self growing base designed to avoid drift and mismatches, instead of a though-before sieve.
We could still wonder why people feel comfortable in a local library, where all documents are well-ordered according to peculiar local concepts, but would feel a lack of awareness in a standardized library environment ?
On the one hand, we can weight the scheme of classification itself. In a sharp-specialized context, users apply to the classification some refinements that they feel pertinent, but which could not match with the formal (e.g. Decimal) classification, and this one is perhaps not fully understood, nor is it necessarily fully approved.
On the second hand, as nested vocabularies apply to given scopes of activity, the dialogue with people of these domains could be improved if acquainted terms were used. Therefore, it seems interesting to make the system able to adapt the vocabulary with the webuser. Words from this vocabulary should be used in the interface screen for labeling the form fields, and to give textual metadata as well.

A crucial step for thematic search : choosing the entry point

The first step in entering an e-library interface is to select an element in the entry table. Several ways are possible. The ones like Grisemine[5] propose to select first the "document family", others, like Castore[6] and many others present a list of scopes, but the selection of a class based on both a thematic class and a type of document is also recommended for example by the Deutsche Initiative für Netzwerkinformation (DINI):
"A rough textual classification facilitates technical structuring and selection. The structure below complies with the functional groups of the German National Bibliography (Deutschen Nationalbibliographie DNB) of the Die Deutsche Bibliothek . Each element should be defined via its functional group number. Descriptions should be output in English (<setName>). An identification via the number will make it easier to convert into the Dewey Decimal Classification (DDC) at a later date at least as soon as the Die Deutsche Bibliothek has converted into the DDC as well."[7]
A priori, the principle adopted by <Document Libre> could remind the DINI recommendations, and this suggested scheme could take place in the <Document Libre> database without any difficulty. Nevertheless, differences exist in that the main concern of <Document Libre> is less to match closely with a given classification scheme, than implementing a sieve able to cluster the entries so as scope and adapted vocabulary match. In other words, we suggest a usage-based classification.
In a rational approach, this attempt should be preceded by a huge semantic study which could define the smart and clear way in clustering scopes.
We approached the challenge in a more realistic process, starting from a generic scheme and staking on a dynamic growth. As we will see below, it will be made use of practices of different kinds of users whenever they are given directly by suggestions or gradually capitalized from the records of the interactions between the system and the webusers.

Entry points : stores and themes

The context is designed according to the image used above, dealing with a customer in a usual library. Terms have been chosen on purpose. Selecting a store leads to the first rough choice. In the very beginning, seven classes have been implemented as "stores" : teaching and research ; literature ; technological documentation and data sheets ; vulgarization and everyday life ; citizen and social activities ; arts , sports and spare time ; bank of images. This first step is a manner of arrowing the path. Just after, a "theme" must be selected. The latter is the actual cluster - the real thematic entry point - whereas the "store" is the title of a group of themes. Since themes are more than hundred, selecting first a "store" limits the sub-item number to about sixteen themes.

Beyond the entry point

Vocabulary fitting

A convenient effect of having selected the entry point is to have selected at the same time the vocabulary tables which best fit to the theme. These tables are intended :

- for labeling form fields in the interface screens
- for recording already-used keywords in this theme.

In addition, all records of nouns and expressions given by users when submitting a document or managing the classification will be clustered specifically to be associated with the theme in the future. Eventually, these records will turn in recommended vocabulary for new submissions. Library keepers which have a privileged access to the database are able at any time to amend these records in order to reorganize or facilitate the convergence.

Bookshelves as thematic nodes

We would introduce the key principle retained to facilitate the indexation along whith thematic nodes relying on the user classes, these classes being typically based on the purpose of the document as well as the supposed level of knowledge of the reader. This approach matches with the instinctive know-how of library assistants who put together the books intended for a given public, on a given subject and coverage. This leads to consider that the most efficient way to express the thematic situation of a document could be a compound of three axes along the topic, the coverage and the required level of knowledge. Related Dublin Core element "Subject", "Coverage" and "Audience" should be the right crossing dimensions.

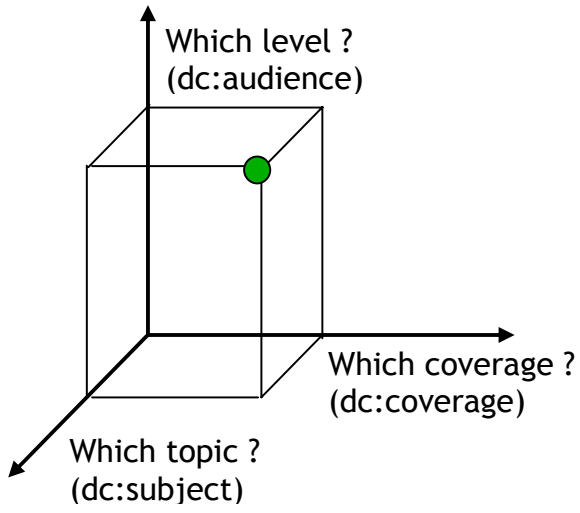


Figure 1 : bookshelf : a three-axial thematic node

Obviously, the matrix built along the full three-axial space of known values of both elements would lead to a tremendous number of nodes. A rather small share of nodes should be relevant.

Let us take an example : technical data sheets do not refer to the whole set of thematic scopes, and the typical definition of coverage[8] does not apply to such resources which are neither concerned by geographical areas nor by dates.

The resolve power of the crossed metadata would be cut down for each domain which is not concerned by the typical definition of coverage. This is the case each time geographic zones, dates or jurisdiction are both irrelevant. This argues for new refines of the coverage element that could apply to other kinds of coverage. For example, "Sensors" are drawn to cover different physical scales : force, pressure, strain... The main advantage of this extension would be keeping the same metadata structure and processing methods for all themes.

Bookshelves as textual metadata

Textual metadata are words or sentences able to describe explicitly the document which they are related to. Terms used to name stores and themes can be considered as textual metadata. In the three-axial space discussed above, each node considered as bookshelf is given a title as explicit as "*Intellectual property in US - Master, Ph.D.*" which represents a textual reminder of the attributes along both axes. These titles are

on the one hand an essential guide for the users. They are means for them to feel like in a traditional library. On the other hand, they become also a helpful guide for inserting new documents without breaking the former order.
Indeed, the bookshelf name cannot replace of a DC element. Subject, coverage and audience should have their proper content. Consequently, bookshelf title can be seen as redundant. But since a kind of fail indicator is required to avoid classification mismatches, this redundancy will probably be found very convenient in practice.

Development of specialized vocabulary
Specialized vocabulary is divided in two classes whether nouns and expressions are "thought before" or are recorded following the use of the system. Several implementations have been made in order to improve :
a) the scheme along which clustering scopes is processed
b) the vocabulary used in the interface screen for labeling the form fields
c) the textual metadata typically used to name bookshelves
d) the set of recommended keywords
Points (a) and (b) lead to static vocabulary ; (b) and (c) lead to dynamic vocabulary.

Development of static vocabulary
As stated before, a starting scheme has been adopted as first entries for the thematic search system. Whenever a whole set entry points could be established after an assessment of typical user groups, many clusters could remain empty for a long while ! It is probably more realistic to build a flexible set based on existing resources, keeping in mind that this set could be extended gradually provided the indexing system is not hard-coded but has been thought to allow convenient and quick modifications. Nevertheless, a starting grid has been proposed. The choices have been made carefully, but are drawn to be amended continuously. Also, it has been thought that specialists from anywhere could bring their own contribution - according to the usages of free software developing process. As a body, this kind of development community should bring the required expertise to amend efficiently the initial scheme. Typically, this collaborative organization should involve library keepers of the sites aggregated in the repository network.
In order to collect suggestions, a *blank table*[9] has been set with first entries listed whereas all other cells have been blanked. It is available on the <Document Libre> site.

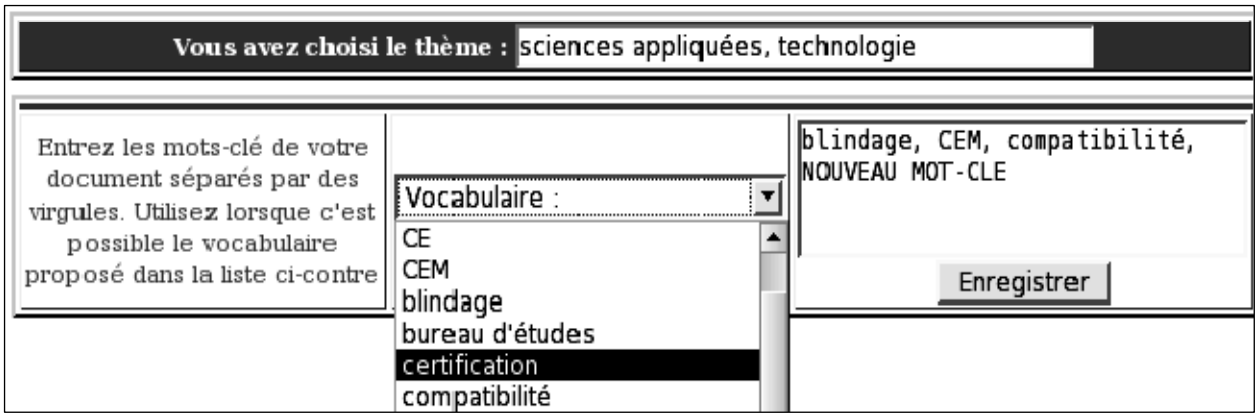
Table 1: excerpt of suggestions table (blank table)

rayon	*theme*	*sujet*	*champ d'application (espace, temps, juridiction...)*	*public de destination*
Enseignement-recherche	informatique, systèmes information			
16 thèmes	philosophie psychologie			
	religion théologie			
	religions			
	sciences sociales			
	sciences économiques			
	sciences juridiques			
	sciences de l\'édudation			
	mathématiques et sciences exactes			
	sciences naturelles			
	sciences médicales			
	sciences appliquées,technologie			
	gestion comptabilité			
	linguistique, philologie			
	hist géo, archéologie géologie			
	Autres thèmes			

Development of dynamic vocabulary
Dynamic vocabulary develops from records of nouns and expressions entered with submissions of new documents.
Each term entered is potentially a metadatum to be. The same words would probably not be used for characterizing on the same "subject" both a scholar work and also a popular work, or for characterizing documents of different types (e.g. a photography vs a non fictional prose. The best vocabulary is then theme dependent.

At the time of submitting their offer, authors are induced to fill a submission form (in three successive screens). Certain nouns must be typed directly, for example the author's name, others are chosen from a popup list, for example the access type ("direct" or "via" access for downloading the document) but in most cases, authors have to choose between using an already recorded word or expression - a suitable set extracted specifically from the database for a given form field being presented in a popup list - or suggesting a new one in a text window. According to this principle, the classification could be suggested by author but this is not mandatory. The decision in matter of classification can be committed to the library keeper.

Figure 2 : Selecting already recorded keywords ans suggesting new ones

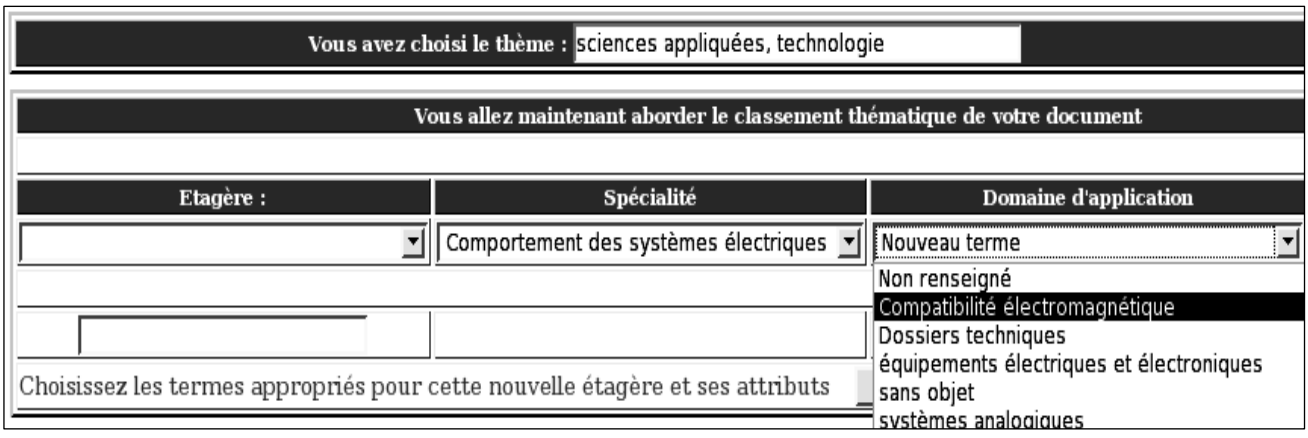


Submission records are XML translated and stored temporarily on the indexation server. An automated e-mail is then sent to the library keeper of the site to which the submission applies, with an included hyperlink that shows him the screen intended for reviewers. He is then able to correct or accept new records that complete the database as soon as the accept clic is entered.

Maintaining vocabulary tables

A comprehensive check up of the vocabulary database is possible and should rely on a collaborative work. The database can easily be managed via phpPgAdmin which is a free software tool allowing an on-line interface to the database tables.

Figure 3 : Selecting already recorded labels and suggesting new ones



Since it is not worth allowing several people to tinker the database, the latter might be modified at once after concerted decisions. Representatives of repositories aggregated to the <Document Libre> network will be welcome to form a development team. Improvements can be experienced from local implementations for research purpose. Since the whole system is based on free software[10], local implementations are free of charge. Moreover, mirroring the database is easy via an export file in text format. Pgsql format is native, XML could be performed as well.

theme	cderayon	tspec	tcl	tannee
Arts plastiques , photographie	2	Activité	Genre	Public destinataire
Cinema, video	2	Thème	Genre	Public destinataire
Collections, généalogie	2	Centre d'intérêt	Epoque, lieu	Public concerné
Gastronomie, art de la table	2	Sujet	Lieu, époque	Public concerné
Jardin, art floral	2	Activité	Espèce	Public concerné
Jeux de société, jeux interactifs	2	Jeu	Catégorie	Public destinataire
La maison	2	Activité	Domaine d'application	Niveau de compétence
Mode et décoration	2	Activité	Domaine d'application	Niveau
Modélisme	2	Activité	Catégorie	Niveau
Musique	2	Activité	Instrument ou Genre	Niveau
Nos animaux	2	Sujet	Espèce	Public concerné
Sports	2	Sport	Type ou Pratique	Niveau
Vehicules et bateaux	2	Activité	Appliqué à :	Expérience

Figure 4 : Online interface to the database

Implementation of the indexing system

In order to clarify the framework of the indexing system, we have first to precise few key points about what is aimed.

- The Indexing system is by no means a library in itself. It offers an indexation service available for independent e-libraries which tend to share their content without being merged in a depersonalized context.
- The share of documents must be done without impacting negatively the role of public relations that a given site is supposed to fulfill. Home page is the normal step to reach a site, and documents are not to be extracted by deep links if this way is not accepted.
- The catalogue of a given e-library and the extent from external repositories are distinguished. An intentional action must be done to extend from the local to the shared content.
- Setting up and managing an e-library or a document collection should be a light work which requires a minimum knowledge and investment. Typically, a basic Internet site and a computer with an e-mail and Internet browser should be sufficient.
- Each indexed file is tracked day by day, and any revision leads to an automated e-mail to the author with an already filled form in order to record metadata modifications if any.
- The system is intended to process a set of requests which return directly as XML or html files the metadata related to a document of a given identifier, or other data like the bibliography related to an author, the result of a search...

As a whole, a given e-library using the system is based on three pillars

- The repository : server on which a group of authors has a writing access and can replace their files if necessary ;
- the e-library : piece of a given Internet site (institutional site... school site...) ;
- the indexation service itself.

Though its address vanishes from library sites[11] the indexing server performs most tasks in the document search and delivery. Typically, the e-library site must record very short html files such as system calls, one or more personalized styles sheets (css type), perhaps a certain number if images, and a short configuration file. An installation kit is available to make the setup as simple as possible. A ready-to-use standard interface can be chosen. If not, a special tool[12] helps in designing personalized screens.

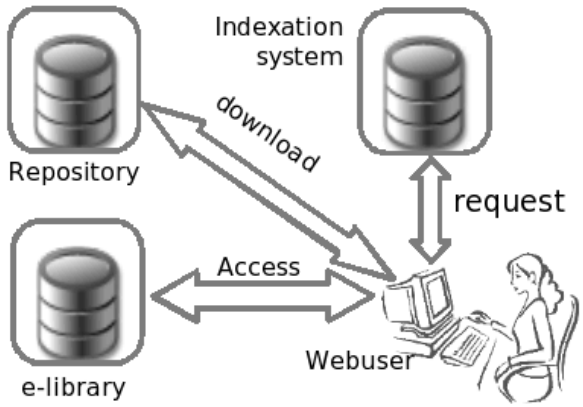


Figure 5 :a Three pillars based system

The indexing system composes html pages at once to fit the requirements of any client along the parameters following the system calls.
The full description of these facilities would be beyond the scope of this paper. We can therefore give below the functional schematic graph of the typical composition process:

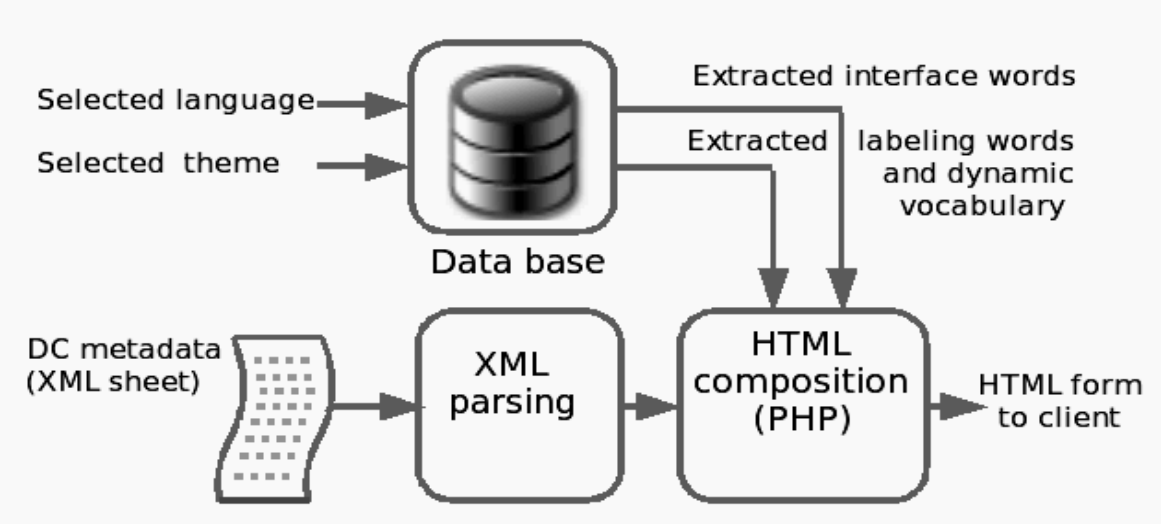


Figure 6: Typical composition process

Conclusion
A breakthrough in the promotion of Grey literature is possible with the concomitant effect of powerful communication, judicious rules in a straight framework , convenient processing tools, good human organization. The Web brought the first requirement, standard metadata brought about the second, we hope that improving indexing tools will help provided a sufficient number of protagonists get acquainted with the possibilities and agree together in a way to gather new sharing practices. The ball is now in their court…

References

1 http://www.documentlibre.org/kit.html the site <Document Libre> is sponsored by the university of Poitiers and is implemented with the support of i-medias

2 Ranger, S.L. - Grey Literature in Special Libraries: Access and Use. - GL6 Conference TextRelease, 2005

3 http://www.iut2.upmf-grenoble.fr/RI3/Lille97_redoc.htm Vers un serveur d'indexation et de recherche sur le Web (1997) Gilbert EYMARD & Jean-Marc FRANCONY Groupe RI3 Université Pierre Mendès France

4 http://debora.enssib.fr/del63.rtf Metadonnées et bibliothèque numérique Youcef Amerouali, 1999 Université Claude Bernard (Lyon 1)

5 http://cridoc.univ-lille1.fr/default.asp?bustl/grisemine, see also : Grisemine, une bibliothèque numérique de littérature grise, Marie-France Claerebout, Jean-Bernard Marino, GL5, dec 2003

6 http://www.emn.fr/x-info/bn/index.html see also : A study of teachers and researchers practices with digital documents - grey or not Author's Name(s): Céline Bourasseau, Cédric Dumas Organization(s): EMN (Ecole des Mines de Nantes) - GL6 Conference Proceedings. TextRelease, 2005

7 Electronic Publishing in Higher Education How to design OAI interfaces - Recommendations - for Data-Provider at german Universities - Deutsche Initiative für Netzwerkinformation e. V. Juillet 2003

8 "Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity)."
see : http://www.dublincore.org/documents/dcmi-terms/

9 Http://www.documentlibre.org/blanktable.html table intended for collecting suggestions

10 OS (Linux), database, page processors are under GNU/GPL or similar licence, the server(Apache) is under Apache licence whenever the indexing system itself is under the CECILL licence.

11 Framing technique allows to make the real server address not visible in the address bar

12 Http://www.documentlibre.org/svg2css.html using OpenOffice.org draw for designing an interface screen.

Author and Title Index 2007

Author and Title Index 2007

About the Authors

Nikkia Anderson is a Technical Information Specialist for Information International Associates, Inc. (IIa). She earned a BS in Computer Science from Bowie State University in Bowie, MD in 2003 and is currently pursuing a MLIS at Drexel University. She is experienced in metadata, taxonomy, and repository development. Other areas of interest include scanning, optical character recognition, and preservation. Email: nanderson@iiaweb.com

Anne Asserson holds a Cand. polit. with a Masters in Information Science from the University of Bergen, UiB. She has been working with Research Documentation, and has participated in substantial parts of CRIS (Current Research Information Systems) developmental work, locally and nationally. Anne led the project establishing and implementing a Research Documentation system, Fdok www.ub.uib.no/fdok/sok, at UiB. he has published in the areas of CRIS and Grey Literature and served on program committees for the GL conference series. Email: anne.asserson@fa.uib.no

Todd A. Chavez is Director of Collection Analysis for the University of South Florida Libraries. His research agenda includes assessment of collections advancing interdisciplinary research and Internet-based subject/discipline knowledge portals. Current projects include the Karst Information Portal and a medieval studies portal connecting Florida's emerging community of medievalists. Email: tchavez@lib.usf.edu

Gail Hodge, IIa Senior Information Scientist, is an authority in the field of information science, with an emphasis on metadata, taxonomy, and thesaurus development. For over 25 years, she has helped organizations develop systems for capturing metadata and for using classification schemes, thesauri, and other knowledge organization sources in a variety of fields, including the environment, life sciences, and aerospace. Email: gailhodge@aol.com

Andrea Japzon is currently pursuing a doctoral degree in information science at Drexel University. She has a BA in English from the University of Florida, MLS from Florida State University and MA in geography from Hunter College of the City University of New York. She has worked for The New York Public Library, the Hunter College Library, the Enoch Pratt Free Library, and the NASA Goddard Space Flight Center Library. She was the recipient of the Society of Woman Geographers fellowship for studies in urban geography and her current studies are supported by an IMLS fellowship. Recent publications include "A neighborhood analysis of public library use in New York city," in Library Quarterly. She has been an active member of ALA for over 12 years and positions held include Chair of the Notable Books Council. Email: ajapzon@pop200.gsfc.nasa.gov

Keith Jeffery is currently Director, IT and International Strategy of STFC (Science and Technology Facilities Council), based at Rutherford Appleton Laboratory in UK. Previously he was Head of Business and Information Technology Department with a staff of 140 supporting over 360000 users, developing software for business and science and doing leading edge R&D. STFC hosts the UK and Ireland Office of W3C and develops and supports the largest OA (Open Access) institutional repository 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

Anna H. Perrault is Professor in the School of Library and Information Science at the University of South Florida. Her bibliometric studies of aggregated academic library collections have won numerous research awards. Collection assessment projects have resulted in legislative appropriations to enhance academic library collections. Email: Perrault@cas.usf.edu

Pete Reehling is GIS & Data Formats Librarian for the University of South Florida Libraries. Currently, he is establishing an Enterprise Geodatabase using Oracle and ArcSDE to create a central GIS repository and evaluate its effectiveness in providing GIS data to the USF GIS user community, including karst scientists. reehling@lib.usf.edu

Debra Revere, MLIS, MA is Lead Research Scientist of the myPublicHealth Project at the Center of Excellence in Public Health Informatics and holds a Clinical Faculty position in the School of Public Health and Community Medicine at the University of Washington, Seattle, WA, USA. She also leads the Telemakus Project in the Department of Medical Education and Biomedical Informatics. Debra came to the UW in 1999 as a National Library of Medicine (NLM) Fellow in the Integrated Advanced Information Management Systems (IAIMS) Program. She has since worked on a number of biomedical informatics projects. In 2003 she received a Visiting Researcher appointment at NLM's Lister Hill National Center for Biomedical Communications to conduct research on applying NLM tools to automate text analysis and indexing of unstructured documents. Her research interests include examining the utility, application, and potential of text mining and indexing programs and spatial-semantic navigation tools for information seeking and retrieval within unstructured document collections. Email: drevere@u.washington.edu

Janet Webster is the head librarian at Oregon State University's Guin Library located at the Hatfield Marine Science Center in Newport, Oregon. This library focuses on marine science and supports an active field research center. Her various interests include collaborative filtering, integration of grey literature into library collections, and global access to local information. Most recently, during her 2004 sabbatical, Professor Webster developed a plan for an institutional repository at OSU and then collaborated with the Fisheries Department of the UN Food and Agriculture Organization to study enhancing access to information that supports implementation of the FAO Code of Conduct for Responsible Fisheries. She is an active member of the International Association of Marine and Aquatic Libraries and Information Centers having served as its president and on numerous committees. She received her undergraduate degree from the University of Chicago and her MLS from Columbia University. Email Janet.webster@oregonstate.edu

https://doi.org/10.26069/greynet-2026-000.608-tgj

Notes for Contributors

Non-Exclusive Rights Agreement

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

Submission Requirements

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

REFERENCE GUIDELINE

General

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

Specific

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

Examples:

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

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

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

Review Process

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

Journal Publication and Article Deposit

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

https://doi.org/10.26069/greynet-2026-000.608-tgj
TGJ Volume 3, Number 3 2007

An International Journal on Grey Literature

'MAPPING GREY RESOURCES'

AUTUMN 2007 – TGJ VOLUME 3, NUMBER 3