HARNESSING THE POWER OF GREY

EIGHTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

Lindy C. Boggs
International Conference Center
New Orleans, Louisiana USA
4-5 December 2006

Conference Proceedings

ISSN 1386-2316

Conference Sponsors

EBSCO
Information Services
United States

NYAM
New York
Academy of Medicine
United States

NLE
National Library
of Education
United States

IIA
Information
International
Associates, Inc.
United States

INIST-CNRS
Institute for Scientific
Technical Information
France

Program Committee

CNR	National Research Council	Italy
euroCRIS	Current Research Information Systems	Norway
GreyNet	Grey Literature Network Service	Netherlands
INIST-CNRS	Institute for Scientific and Technical Information	France
JST	Japan Science and Technology Agency	Japan
NYAM	New York Academy of Medicine	USA
NYU	New York University School of Medicine	USA
OSTI-DOE	Office of Scientific and Technical Information	USA
UCI	University of California, Irvine	USA

Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
www.textrelease.com · conference@textrelease.com
T/F +31 (0) 20-331.2420

CIP

GL8 Conference Proceedings
Eighth International Conference on Grey Literature : Harnessing the Power of Grey, 4-5 December 2006 / compiled by D.J. Farace and J. Frantzen ; GreyNet, Grey Literature Network Service. – Amsterdam : TextRelease, February 2007. – 132 p. ; 30 cm. – Author Index. – (GL-conference series, ISSN 1385-2316 ; No. 8)

INIST-CNRS, NLE, and NYAM are corporate authors and associate members of GreyNet. The plenary and parallel sessions contain the full text of fifteen papers presented during the two-days of conference sessions. The print copy has been standardized in font, size and format for general layout and appearance. Included is a List of Participating Organizations and Author Index, as well as information on the authors.

ISBN 90-77484-08-6

Foreword

“Harnessing the Power of Grey”

The Eighth International Conference on Grey Literature provided a global platform situated and constructed for the R&D community. One that was focused on the state of the art in grey literature with applications and innovative uses in and for science and technology. The past three conferences in the GL-Series brought to the forefront grey literature in networked environments, works-in-progress, and open access to resources. GL8 must now harness all of this in an effort to demonstrate the power of grey to other information professionals as well as policy and decision makers, funding bodies and new investors.

GL8 offered a solid platform in a metropolitan city devastated by Hurricane Katrina. This city and coastal region is now in the throws of reconstruction unprecedented in recent history. The grey literature community was offered a tiered challenge last December in New Orleans. They were asked to demonstrate the state of the art in their field. They were encouraged to incorporate new and emerging areas in grey S&T. And, they were charged to address echelons within their own organizations responsible for the appropriation of material and human resources.

“Harnessing the Power of Grey” is outlined here by four main themes dealing with

- Collection Development, Collection Policies, and Collection Rescue
- Metadata schemes, Repositories and Software, Standards and Quality Assessment
- Economic and Legal Aspects of Grey
- Mapping Grey Resources for Coastal and Aquatic Environments

GL8 provided the R&D community with a variety of settings in the presentation of their research results. On behalf of the Program Committee and the Session Chairs, I take this opportunity to thank those who provided a content contribution to the Eighth International Conference on Grey Literature. I also welcome those reading these proceedings to voice their comments and recommendations either directly to the authors or via GreyNet.

Dr. Dominic J. Farace
Grey Literature Network Service

Amsterdam,
February 2007

GL8 Conference Sponsors

EBSCO, USA
EBSCO Information Services

NYAM, USA
The New York Academy of Medicine

NLE, USA
National Library of Education

IIA, USA
Information International Associates

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

GL8 Program Committee Members and Organizations

Lily Pregill and Lea Myohanen *(Co-Chair)*
The New York Academy of Medicine, USA

Anne Asserson
euroCRIS, Current Research Information Systems, Norway

Marcus A. Banks
New York University School of Medicine, USA

Deborah Cutler
Office of Scientific & Technical Information, Dept. of Energy USA

Dr. Dominic J. Farace
Grey Literature Network Service, Netherlands

Julia Gelfand
University of California, Irvine, USA

Dr. Suguru Ishiguro
Japan Science and Technology Agency, Japan

Daniela Luzi
National Research Council, Italy

Dr. Joachim Schöpfel
Institut de l'Information Scientifique et Technique, France

Contents

OPENING SESSION

Chair: Marcus A. Banks, New York University, School of Medicine (USA)

SESSION ONE
COLLECTION DEVELOPMENT, COLLECTION POLICIES, AND COLLECTION RESCUE

Chair: Joachim Schöpfel, INIST-CNRS (France)

SESSION TWO
METADATA SCHEMES, REPOSITORIES, SOFTWARE, AND STANDARDS

Chair: Gina R. Costello, Louisiana State University (United States)

SESSION THREE
ECONOMIC AND LEGAL ASPECTS OF GREY

Chair: Anne Asserson, University of Bergen (Norway)

SESSION FOUR
MAPPING GREY RESOURCES FOR COASTAL AND AQUATIC ENVIRONMENTS

Chair: Anne Asserson, University of Bergen (Norway)

PARALLEL SESSION ONE

Chair: Gretta Siegel, Portland State University (USA)

PARALLEL SESSION TWO

Chair: Bertrum H. MacDonald, Dalhousie University (Canada)

Publicis Nancy - 03/05

SCIENTIFIC AND TECHNICAL DOCUMENT DELIVERY

Find and receive the information you need in a simple and easy way.

With INIST-CNRS, the leading French scientific document delivery center, you are certain to obtain a copy of the majority of the documents you need in Science, Technology, Medicine, Humanities, and Social Sciences.

AN EXCEPTIONAL MULTIDISCIPLINARY DOCUMENT COLLECTION

The collections of INIST and its international network of partner libraries cover the core literature published worldwide in Science and Technology.
INIST's own collection contains 26 000 French and international periodical titles,including 8 000 current subscriptions, as well as a significant number of scientific reports, conference proceedings and doctoral dissertations.

ORDERING AND DELIVERY, TAILOR-MADE SERVICES

INIST offers a complete range of tailor-made services for you to locate, order and receive copies of the documents you need.

If you are an occasional user, the ArticleSciences search engine is available to order an article and pay for it online with a bank card.If you are a frequent user, after opening an account at INIST, you can use an interface (http://services.inist.fr) that offers a whole range of ordering and payment methods tailored to your needs.

For more information: **http://docdelivery.inist.fr**

inist INSTITUT DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE

Contact and information: INIST-CNRS - Customer Services
2, allée du Parc de Brabois - F-54514 Vandœuvre-lès-Nancy Cedex
Téléphone : +33 (0)3 83 50 46 64 – Télécopie : +33 (0)3 83 50 46 66
Email: infoclient@inist.fr

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

GL8 Welcome Address

Fredrick Barton
Provost and Vice Chancellor,
University of New Orleans, United States

On behalf of the University of New Orleans and Chancellor Tim Ryan, on behalf of the City of New Orleans and the State of Louisiana, and for those of you who have traveled here from abroad, on behalf of the United States, I welcome you to the Lindy Boggs Conference Center and wish to all of you a productive meeting and a pleasant visit to our campus and our community.

One hundred fifty years ago, when cotton was the king of international trade and, drawing upon the great commercial highway of the Mississippi River, New Orleans was king of cotton, our city was the richest in the world, as the magnificent homes in the Garden District and along Esplanade Avenue survive to testify.

The passage of time and the evolution of world economies have diminished New Orleans as a central place of material wealth. But we remain almost astonishingly culturally rich. Our music and our cuisine are exported to and emulated around the world. Louis Armstrong and Wynton Marsalis learned to play their trumpets here and have gone on to perform for audiences worldwide. Emeril Lagasse, Paul Prudhomme and Frank Brigtsen are household names for those knowledgeable about gourmet cooking and fine dining.

New Orleans is a city of festivals, a city that likes to dress up and, as our students like to say, also really likes to get down. It is a city that relishes a celebration and, more than any other in America, loves a parade. Millions of people have visited New Orleans for Mardi Gras and for the Jazz and Heritage Festival. But these are not tourist activities, for the greatest devotees of these celebrations are those who were born here, live here and raise their children here.

New Orleans is a city of literary lights. Two of our festivals celebrate the works and legacies of William Faulkner and Tennessee Williams, both of whom lived here and wrote major works in these environs.

A great city needs a great public university, and for the last half century the University of New Orleans has endeavored to fill that role. UNO was founded in 1958, among the oldest post-World-War-II urban universities that responded to the requirements the economies of the 20th century vested in higher education. UNO opened its doors on September 5, 1958, as an undergraduate college teaching 1,500 students that first year, twice as many as were expected. We were the first racially integrated public university anywhere in the American South, our proud tradition of cultural diversity present in the first moments of our history Today, UNO has evolved into a comprehensive research university with scores of degree programs at the bachelor's, master's and doctoral levels. UNO's students come from across the country and from nearly 100 foreign countries. In addition to our programs on campus, we operate international programs in five European countries, in the West Indies, in Latin American and in Asia.

Unfortunately, as you will learn during your visit, in the aftermath of Hurricane Katrina, the city of New Orleans and its great public university that is UNO both face huge challenges. The city has lost housing for 300,000 people and that means that 60% of the city's residents have relocated outside its boundaries. The metropolitan area has lost a third of its citizens to relocation outside the area. Meanwhile, UNO has lost a third of its students along with the missing core population base. The lost tuition revenue from the displaced students has placed a severe strain on the university's ability to operate.

The sorrow that the remaining residents face is registered in the sagging houses you will see on so many streets as you make your way to this center during your conference. In neighborhood after neighborhood, homes are windowless, abandoned, forlorn, properties drooping into uselessness like chocolate figurines left too long in the summer's heat. So short a time ago each of these decaying edifices was someone's home where good, spicy food simmered in the kitchen and the laughter of full lives echoed within its walls. Now the air smells of mold and mildew, and inside the walls silence reigns. Ruined lawns, broken sidewalks and snaggled streets breed despair, house to house, block to block

The sorrow we face is registered in every destroyed school, its playing children vanished, in every church where hymns are no longer sung, in every store where goods are no longer sold, in every café, restaurant and bistro where our good food is no longer served and where friends near and far no longer gather.

Those of us who work at UNO know this devastation. For these are our houses and those of our colleagues and those of our students and those of our friends. And wherever we now reside, we cannot reach our workplace without driving through flooded neighborhoods, where heartbreak is our daily passenger.

Those of us who hail from New Orleans know something about the great flood of 2005. It was a disaster made not by nature but by man. Hurricane Katrina did not sweep over this city; it broke through to this city. Our levees were high enough, but they were not strong enough. Our homes were lost, our lives were altered, not as an act of God, but as an act of negligence, not as the product of inevitability but as the byproduct of irresponsibility.

The current sad state of New Orleans teaches us immediately that we must do better. We must immerse ourselves in research, and we must employ the research that we undertake. We must have access to it; we must learn from it; we must apply it.

This, in the broadest sense, is what all of you are about and why you are here with your colleagues at the Lindy Boggs Center. Tomorrow, you will have a session of this conference dedicated to issues of coastal erosion. The coast line of this state is disappearing at a terrifying rate. Among the results of our evanesced wetlands is the increased reach of hurricanes such as Katrina. Once, stubborn grasses and other swamp vegetation stood as an obstacle to the great storms roaring out the gulf. Land now missing stripped the wind of its power and slowed the force of the tidal surge. But where once hundreds of miles of wetlands grew, now stands only open water, a fast track right to the heart of our city. Learning about coastal erosion, preserving what we have learned and can learn, expediting access to that knowledge, applying that knowledge as buffer to any repeat of what we have recently endured: this is not just our opportunity; this is our duty. And applied writ large to the many crises humankind faces and must face down, this is what you do and why what you do is so very important.

I wish you much success in that session on coastal erosion and in all that you deliberate about here in our conference city. I hope that you benefit from your time here together. And I hope that we benefit from what you bring and take away from here.

My heartiest welcome and best wishes to you all.

Grey Literature: Taxonomies and Structures for Collection Development

Julia Gelfand
University of California, Irvine Libraries, United States

Abstract
Libraries worldwide have not picked up the pace of addressing where Grey Literature fits in collection policies. This remains rather curious due to the skyrocketing prices of traditional books, journals, databases, and other information resources, and trends to serialize and promote access in perpetuity. Most collection development policies only address resources for which payment has been made, where formal acquisitions or licensing practices are observed. Due to more interest and a commitment to Open Access initiatives and electronic publishing, Grey Literature does not appear to have a more stable and comfortable home in libraries, although it has demonstrated increasingly how it is being cited more seriously and frequently. Often, content that is openly available on the Internet and for which there is no required payment finds no bibliographic control or metadata associated with it that begs for description and order. Thus, this paper will examine what kind of alternatives there are for discovering, cataloging and processing the immense grey literature so that additional value and access is guaranteed giving it credibility in a collection development policy.

Building on the celebrated works of Edward Tufte's, Envisioning Information (1990), Davenport and Prusak's, Information Ecology: Mastering the Information and Knowledge Environment (1997), and the pioneering work of Bloom and Krathwohl's, Taxonomy of Educational Objectives (1956), the core component of information architecture suggests how taxonomies are a foundation for visual design of information navigation and structurally define relationships of different elements in a cohesive package. Several key examples of Grey Literature in the Social Sciences, Arts and Scientific disciplines and in Digital Libraries will be used to demonstrate how a taxonomy contributes to the outline of most Collection Development policies and establishes relationships by format, organization, finding tools, and access points. Policies are what drive and determine what libraries acquire and license, point to and promote in their catalogs by an increasingly important web presence. Grey Literature needs to share more equal billing in terms of discovery and retention and unless it is included in the formal collection development policies the added value of incorporation is weakened. If information usage patterns are indeed more reflective of information architecture, then the taxonomy structure should encourage collection development policies to entertain more Grey Literature content. This paper will consider how libraries would benefit by such recommendations and become more relevant to its users. Illustrated comparisons of what new roles a library would experience with more Grey Literature referenced and alluded to in its collection policy will enhance the role of bibliographers and invite more widespread global content with less financial demand than other information products.

Organizing information has been the hallmark and mainstay of libraries for their entire history. In traditional print collections, classification systems were introduced to organize material by subject and the most widely adopted method in large academic or research libraries has been the Library of Congress Classification scheme, or LCSH. This alpha-numeric assignment of subjects is probably the most common worldwide scheme of the modern library because it can be expanded with ease to incorporate greater specificity. The most basic subject outline is this with follow-up breakouts of up to three letters and four numbers to the right of the decimal point:

- A -- GENERAL WORKS
- B -- PHILOSOPHY. PSYCHOLOGY. RELIGION
- C -- AUXILIARY SCIENCES OF HISTORY
- D -- HISTORY (GENERAL) AND HISTORY OF EUROPE
- E -- HISTORY: AMERICA
- F -- HISTORY: AMERICA
- G -- GEOGRAPHY. ANTHROPOLOGY. RECREATION
- H -- SOCIAL SCIENCES
- J -- POLITICAL SCIENCE
- K -- LAW
- L -- EDUCATION
- M -- MUSIC AND BOOKS ON MUSIC
- N -- FINE ARTS
- P -- LANGUAGE AND LITERATURE

- Q -- SCIENCE
- R -- MEDICINE
- S -- AGRICULTURE
- T -- TECHNOLOGY
- U -- MILITARY SCIENCE
- V -- NAVAL SCIENCE
- Z -- BIBLIOGRAPHY. LIBRARY SCIENCE. INFORMATION RESOURCES (GENERAL) [i]

Two choices usually exist when cataloging information that pertains to grey literature in particular. Assignments will be made within the subject structure or assigned to either T10.7 which describes "Communication of Technical Information" or to the ZA subclass for Information Resources with a breakout resembling:

ZA3038-5190	Information resources (General)
ZA3150-3159	Information services, Information centers
ZA3201-3250	Information superhighway
	Information resources (General) - Continued
ZA4050-4775	Information in specific formats or media
ZA4050-4480	Electronic information resources
ZA4150-4380	Computer network resources
ZA4450-4460	Databases
ZA4550-4575	Motion pictures. Video recordings
ZA4650-4675	Pictures. Photographs
ZA4750-4775	Sound recordings
ZA5049-5190	Government information

One can easily see how inflexible this structure becomes because browsing by location either in a collection or online is increasingly rigid and is determined by the subject headings assigned to the item record. The library practice of classification is the process of analyzing content and determining where in a taxonomy it belongs results in the assignment of metadata tags. When one looks for "grey literature" in a Union Catalog one sees many entries corresponding to a range of classifications, but in reality this does not capture the content beyond the mapping of its content into a taxonomic system that can be classified.. Thus, we conclude why the definition of grey literature encompasses its difficulty in acquisition and processing, "Information produced 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*." (Luxembourg, 1997 - Expanded in New York, 2004)"

Gokhale in 1997 probably was the first to write about a taxonomy to describe grey literature although she did describe it as such. She introduces a Generic Design Science methodology which identifies:

- Forms
- Disciplines
- Usefulness to different constituencies
- Availability
- How it is Managed

as categories to qualify its utility.[ii] This is done by qualifying each example of a grey literature source with the specific attributes. Far from sophisticated, it offers additional matching to content that may not ever have experienced full cataloging or processing and does not include examples of digital content although it would be easier to create a database with links for that format.

My earlier work from the 1990s explored the dichotomies in library collection development efforts and policies and then Siegel followed by Lehman and Webster[iii] examined in different surveys how academic libraries were beginning to treat grey literature in collection development policies and determined that little change was noticeable amongst the grey literature found in academic library collections (operative word is "found"). Interestingly, their work in 2002-2004 paralleled the evolution of the early digital library movements defining their paths and it remains clear that acquisitions and cataloging efforts to support grey literature remain a dilemma for libraries. Last year I explored evidence-based practices and how that served grey literature, and I offer the premise that libraries are indeed capturing grey literature, adding it to collections but it is increasingly mainstream as it falls under the rubric of electronic resources. Statistics from the Association of Research Libraries (ARL) illustrate the large increase in such holdings over the last decade. Exactly five years ago, in December 2001 ARL reported that its member libraries spent more nearly $100Million on electronic resources, an increase of 6.3% over the previous year.[iv] This major increase was due to more mature consortia efforts to invest in big packages of online journals, early generation eBook packages and the select individual resource, that may be indeed a shade of grey. Many

libraries are adding such content even if it is free as long as it is valuable and stable, with an expectation that it will be maintained for however the future is defined. This is particularly noticeable when browsing in library catalogs such as OCLC's WorldCat (http://worldcat.org/) and can be extended to member libraries around the globe. There is less that distinguishes grey literature from other electronic resources. Current trends indicate that specialized institutional repositories are now commonplace at most academic institutions, government and corporate libraries and each tries to capture and retain their own intellectual capital created by local talent and faculty, with support from resources and grants or gifts awarded to the institution, and also to customize the content. Digital library emphases and collection units are inherent in the reshaping of the academic library enterprise and address a wider scope of materials identified, acquired and processed by the Libraries to support instruction and research. Scholarly communication efforts also contribute to promoting authors' rights and encouraging authors to retain copyright and participate in the open source movement prevalent today. Levy and Marshal have identified a triage to define characteristics common to digital libraries:

1. Documents - Contain *fixed permanent documents* with potentially a changing rate of change and varying duration meaning that they can be permanent or transient depending on the format and content or how highly ephemeral it is;
2. Technology - Digital libraries are based on digital technologies but in reality all libraries are quasi-digital with integrations into huge print collections. One is beginning to observe that more of the holdings are digital but the coexisting model continues to be that of a blended *hybrid*.
3. Work - Digital libraries are to be used by individuals working alone - images prevail that both library users and library staff or service providers work independently even though we consider learning in libraries and serving them to be very collaborative experiences. [v]

Taxonomies are schemas or systems for naming and organizing things into groups that show similar characteristics. The *Oxford English Dictionary* (OED) defines taxonomy as a "classification, esp. in relation to its general laws or principles; that department of science, or of a particular science or subject, which consists in or relates to classification; especially the systemic classification of living organisms."[vi] Taxonomies have been used in many fields for a long time. For example, in Botany, theya re used for plant classifications, e-business organizations harness taxonomies such as Google or Yahoo to develop a structure of topics searchable on the web. Within Knowledge Management a taxonomy breaks and defines an organization into different types of functionality but two key concepts describe this utility: 1) to navigate/find and 2) to determine expertise hierarchically in organizations.

Given that broad description, we can see how adopting a taxonomy contributes to accomplishing some of the goals for learning and communication, a vital goal of grey literature which can be described as:

- Refine reasoning abilities
- Develop stronger arguments
- Communicate complex cases
- Produce better documents
- Make better decisions

Interestingly, one of the most interesting explanations of how thesauri, taxonomies and ontologies are used in the library and information science milieu is provided by Gilchrist as he demonstrates the etymological interpretations. He defines and explores what he calls, "triggers" that reinforce why a taxonomic structure may be applicable to different information commodities or products and I urge us to consider this match with grey literature as we further explore the relationships to other content areas in library collections. These triggers include:

1. Information overload - search engines are challenged to effectively handle fulltext coverage of already large databases
2. Information literacy - users are not always effective searchers leading to less than helpful retrieval
3. Organizational terminology - keyword searching has assumed the preferred method of searching over thesauri verification or other methods of controlled vocabularies and descriptors
4. "Destructuring" of organizations - the fast-paced structure of the information industry in recent years suggests confusion and how users co-mingle sources of information[vii]

It is easy to see why museums use taxonomies - they are pathways that put things in order. Gilchrist also lists five examples of how taxonomies can be used by information providers or corporate entities that issue information. I have extended that to suggest how they are also relevant to grey literature, all of which reinforces the power of grey literature, because without identifying it as such, each is a good example in how retaining common attributes of the modern grey literature as we know it today. They are:

1. Web directories - far too many to note but each is a classification system with different access points
2. Taxonomies to support automatic indexing - a database containing a range of products, such as theses and dissertations, standards, tests, terms, etc
3. Taxonomies created by automatic categorization - those that can interpret text and add another dimension to the original layer making it 2D or 3D as with statistical analysis
4. Front end filters - extending query formation so that user can browse up and down a hierarchy or jump to related terms or content - another method of customized navigation that is simple and directed by user's preferences
5. Corporate taxonomies - drug and pharmaceutical directories, inventories of products, or any other large database that offers sorting and mapping to use to its full potential[viii]

Benjamin Bloom, fifty years ago headed a group of educational psychologists who developed a classification of levels of intellectual behavior important in learning. He identified six levels within the cognitive domain, from the simple recall or recognition of facts, as the lowest level, through increasingly more complex and abstract mental levels to the highest order which is classified as evaluation. Thus, the pyramid has these components which has the following verbs associated with each element:

1. Knowledge: arrange, define, duplicate, label, list, memorize, name, order , recognize, relate, recall, repeat, reproduce
2. Comprehension: classify, describe, discuss, explain, express, identify, indicate, locate, recognize, report, restate, review, select, translate
3. Application: apply, choose, demonstrate, dramatize, employ, illustrate, interpret, operate, practice, schedule, sketch, solve, use, write
4. Analysis: analyze, appraise, calculate, categorize, compare, contrast, criticize, differentiate discriminate, distinguish, examine, experiment, question, test
5. Synthesis: arrange, assemble, collect, compose, construct, create, design, develop, formulate, manage, organize, plan, prepare, propose, set up, write
6. Evaluation: appraise, argue, assess, attach, choose, compare, defend, estimate, judge, predict, rate, core, select, support, value, evaluate

What this may mean in a product relationship with a more outcome focus -illustration that is more descriptive of content than learning objectives and relevant to content analysis, may include these descriptors:

1. Knowledge - the remembering or recalling of appropriate, previously learned information: defines, describes, enumerates, identifies, labels, lists, matches, names, reads, records, reproduces, selects, states, views
2. Comprehension - grasping or understanding the meaning of informational materials: classifies, cites, converts, describes, discusses, estimates, explains, generalizes, gives, examples, makes sense out of, paraphrases, restates (in own words) summarizes, traces, understands
3. Application - the use of previously learned information in new and concrete situations to solve problems that have single or best answers: acts, administers, articulates, assesses, charges, collects, computes, constructs, contributes, controls determines, develops, discovers, establishes, extends, implements, includes, informs, instructs, operationalizes, participates, predicts, prepares, preserves, produces, projects, provides, relates, reports, shows, solves, teaches, transfers, uses, utilizes
4. Analysis - the breaking down of informational materials into their component parts, examining such information to develop divergent conclusions by identifying motives or causes, making inferences, and/or finding evidence to support generalizations: breaks down, correlates, diagrams, differentiates, discriminates, distinguishes, focuses, illustrates, infers, limits, outlines, points out, prioritizes, recognizes, separates, subdivides
5. Synthesis - creatively applying prior knowledge and skills to produce a new or original whole: adapts, anticipates, categorizes, collaborates, combines, communicates, compares, compiles, composes, contrasts, creates, designs, devises, expresses, facilitates, formulates, generates, incorporates, individualizes, initiates, integrate, intervenes, models, modifies, negotiates, plans, progresses, rearranges, reconstructs, reinforces, reorganizes, revises, structures, substitutes, validates
6. Evaluation - judging the value of material based on personal values/opinions, resulting in an end product, with a given purpose, without real right or wrong answers: appraises, compares & contrasts, concludes, criticizes, critiques, decides, defends, interprets, judges, justifies, reframes, supports[ix]

In creating new forms of intellectual capital or information products, especially as electronic resources that meet the criteria of grey literature, there are some potential activities and specific products associated with them that have specific attributes and can be translated as follows:

Knowledge

Useful Verbs	Examples of potential activities or new information products
Tell	Chronology, timeline or list of events
Describe	Facts chart
Relate	Oral histories
Name	Recitations or audio treatment

Comprehension

Useful Verbs	Examples of potential activities or new information products
Explain	Images
Translate	Multilingual treatises
Describe	Abstracts or secondary products
Outline	Finding aids or item descriptions (as in archival support)

Application

Useful Verbs	Examples of potential activities or new information products
Solve or construct	Create models or dioramas
Illustrate	Photo albums, scrapbooks, videos, archives
Classify	Market strategy of repurpose information
Show	Offer directions, maps

Analysis

Useful Verbs	Examples of potential activities or new information products
Analyze	Design survey or questionnaire to query subjects
Investigate	Strategy to find or complete puzzle
Compare/contrast	Create flowcharts or graphs
Identify	Genealogy or family tree
Advertise	Create a commercial or ad to describe product or service

Synthesis

Useful Verbs	Examples of potential activities or new information products
Create	New product
Invent	Design, invent or re-engineer & register new product or idea
Imagine	Make something up - ie) new language code, etc
Compose	Accompanying lyrics

Evaluation

Useful Verbs	Examples of potential activities or new information products

Judge	Render decisions - legal criteria or outcomes
Justify	Add clarity to rationale or annual reports
Argue/debate	Texts & treatment of debates & decisions
Verify	Trace path of information query citing sources, bibliography
Assess	Process & method of evaluation[x]

Edward Tufte responded to a big challenge in 1990 when he examined how to categorize charts, diagrams, graphs, tables, guides, instructions, directories and maps - this content may accompany text but can stand on its own. This increasingly proliferating content that can now easily be part of electronic publishing and is best captured or described as images is some of the new forms of grey literature that are really difficult to process without the structure of a database with like content. What Tufte and his colleagues proved was that their data intensive, static and graphic rich content have new ways to be presented and thus organized, searched, and recalled.[xi]

Today, there are literally uncountable examples of this and many sources that do nothing but contain such presented content. They are known as visual resources, and find themselves in websites with increasingly strong and robust search engines and navigation instruments that allow for description and searching. Some random examples are the David Rumsey collection of historical maps (www.davidrumsey.com); museum collections of arts and design such as from MOMA (www.moma.org/collection),or the high-end Mellon funded ARTstor (www.artstor.org/info),where the operative word becomes "collection." These are some examples of concentrated strength in what was formerly considered grey but is now so robust and dependable that it confirms how major collections whether in private hands or in major national collections all have an online presence. They have also achieved this status due to increased power in software with such products as Luna Imaging Software (www.luna-imaging.com/insight/index.html) and other tools now commonly in the marketplace not requiring so much in-house creation.

Shneiderman, a renowned computer scientist and forward thinker, a decade ago submitted a plan that was a task by data type taxonomy with seven data types and seven tasks and this, too, I surmise is relevant to how grey literature can be characterized. The taxonomy was defined for information visualizations which have some parallel attributes to grey literature and other electronic resources. He prefaces his findings by suggesting that "Exploring information collections becomes increasingly difficult as the volume grows. A page of information is easy to explore, but when the information becomes the size of a book, or library, or even larger, it may be difficult to locate known items or to browse to gain an overview."[xii] The Visual Information Seeking Mantra of "Overview first, zoom and filter, then details-on-demand"[xiii] represents how he characterized the multiple information visualization innovations he observed operating within academic, government and industry research environments. The data types were named in accordance to their complexity or usage patterns they required:

1. one-dimensional data
2. two-dimensional data
3. three-dimensional data
4. temporal data
5. multi-dimensional data
6. tree data
7. network data

and the seven tasks:

1. overview - of entire collection
2. zoom - focus on items of interest
3. filter - deflecting uninteresting content
4. details-on-demand - select as needed with appropriate details
5. relate - view relationships among items
6. history - retaining for recall, replay and progressive refinement
7. extract - allowing for extraction of sub-collections and of the query parameters

Shneiderman concludes with:

> "Although the computer contributes to the information explosion, it is potentially he magic lens for finding, sorting, filtering and presenting the relevant items. Search in complex structured documents, graphics, images, sound or video presents grant opportunities for the design of user interfaces and search engines to find the needle in the haystack. The novel-information tools, such as dynamic queries, treemaps, fisheye views, parallel coordinates, starfields and perspective walls, are but a few of the inventions that will have to be tamed and validated."[xiv]

So what does this do for libraries, digital libraries, collection development policies and grey literature. It offers a structure and forces an organization to ponder how indeed different grey literature is from other materials. Many information scientists such as Borgman, Saracevic and Kantor, Covi and Kling and Lesk[xv] have underscored the importance of different strata or content filters in digital libraries, but the most compelling piece is that by Demas, McDonald and Lawrence. This trio writes a decade ago that there are indeed challenges for collection development with Internet resources. If one assumes that grey literature today is mostly released in digital formats, their experience at Cornell University where they created a conceptual model for selection of electronic material is compelling. This taxonomy of "genre of information resources was evolved to categorize information resources according to their characteristics, how they are used, with similarities in systems of access. Information genre include bibliographic, numeric, spatial and geographic, full text, applications software, sound and image."[xvi] The Cornell template became a record of the collection development policy and collecting intensity level for each taxonomic category of Internet resources. This included:

1. Definition /defining characteristics
2. Typical examples
3. Collection policy notes/collection level
4. Selection questions and guidelines
5. A list of selection tools useful for identifying Internet resources

Harder questions were asked about justification and needed online support. This was early web days but there was an effort to determine the extent to which mainstreaming Internet tools and integrating the selection process could be done. The taxonomy had basically fifteen elements - with subparts of which some examples follow:

1.0 Reference resources
    1.1 Directories
    1.2 Dictionaries
And examples of Grey Literature included:
4.0 Discussion Groups
5.0 Numeric Files
6.0 Genetic Information
7.0 Gophers, gateways and networks
8.0 Museum Catalogs
9.0 Software Archives
10.0
11.0 Graphic image archives
12.0 Sound
13.0 Video conferences
14.0 Publications of the US Government Agencies
15.0 Staff use resources
    15.1 Selection tools
        15.1.4 Publishers Catalogs[xvii]

A sample categorization technology hierarchy according to the Delphi Group may include categorization approaches that include:

- **Manual Advantages** such as human judgment, high accuracy and disambiguation
- **Manual Disadvantages** defined as labor intensive, inability to scale and expensive resources
- **Automatic Advantages** - handles huge volumes, scales easily and inexpensive resources
- **Automatic Disadvantages** - Rule/algorithm fragility, fraught with inaccuracies, and difficult to train
- **Hybrid Advantages** - high volume + accuracy = human-guided rule sets, incremental learning is the norm over ease and speed
- **Hybrid Disadvantages** - has management challenges, requires special skills and maintenance effort[xviii]

What we learned from this is that a new information ecology was born. Bonnie Nardi and Vicki O'Day describe that as a

> "system of people, practices, values and technologies in a particular local environment…A Library is an information ecology. IT is a place with books, magazines, tapes, films and librarians who can help you find and use them….In a library, access to information for all clients of the library is a core value. This value shapes the policies around which the library is organized, including those relating to technology. A library is a place where people and technology come together in congenial relations, guided by the values of the library."[xix]

The information ecology web defined by Davenport a few years prior to Nardi and O'Day's work stresses how "Creating change in an information ecology is clearly a complex, multifaceted undertaking. As in a

natural ecosystem, a change in one environment will effect the others."[xx] The emphasis to make is that these definitions and practices come from anthropologists and management specialists, not librarians yet information ecology describes the complex intersections between data, information and knowledge which are not easy to separate and lend to how grey literature was defined.
Libraries were attracted to the notion of evolution via technology yet these authors reject that information is different in an analog world than in a digital environment. The maturity factor is what counts here as illustrated by Davenport:

Data -->	**Information** -->	**Knowledge**
Simple observations of states of the world	Data endowed with relevance and purpose	Valuable information from the human mind - incl. reflection, synthesis, context
Easily structured	Requires unit of analysis	Hard to structure
Easily captured on machines	Need consensus on meaning	Difficult to capture on machines
Often quantified	Human mediation necessary	Often tacit
Easily transferred		Hard to transfer[xxi]

Grey literature has attributes of data, information and knowledge but translating that into a collection policy is rarely done. Why, when electronic publishing becomes increasingly the norm, and libraries are more and more hybrid organizations containing all formats with the fastest growth segment among digital resources, has this not changed? I ask this because the literature search for this paper reveals very little in the last few years has been written about changing collection development policies tracking the content mix of libraries. One can more easily find policies on the web and they all tend to have a section about electronic resources but none contained grey literature as an indexed item. Even the sections on Electronic Resources rarely distinguish between born digital and conversion of formats.

However, elements of grey literature are included in collection development policies as separate products such as for dissertations, spatial information, images, computer games. This often intangible quality of relevance at time of need is very interesting and hard to apply but it means a lot to library users and consumers. In the practice of management consulting that is what determines the future successes of marketing and product development. If library collection development policies are that important they should reflect this information ecology perspective in a taxonomic structure much like Demas, McDonald and Lawrence attempted a decade ago. With many more eResources in libraries, a more robust digital library movement and the paradigm shift of library collections in general, the concept of critical mass emerges and I believe we would find a refined strategy of the Cornell model still a valuable resource tool. Now late in 2006, we may want to tweak Shneiderman, Bloom and Gilchrist into a different pyramid, neither inverted nor reinforcing hierarchy, but demonstrating:

Resource Identification
Function
Relevance
Knowledge

This simplified taxonomy promotes relevance while emphasizing collection development/management and content strategy and suggests how creating a model of this combination will better enhance the classification schema with more description than LCSH. If information ecologies can embrace taxonomic structures, then the impact of electronic information sources on collection development will be more transparent to those who select, process and use the content. Taxonomies will support **categorization** for browsing and search functions, **profiling** to alert users to new relevant sources, **synchronization** to find and update all content accessible to readers in a secure technically robust way, **offer enhancements** to create new knowledge such as through multilingual or special services function. We can conclude that grey literature is format agnostic and the proposed information architecture contributes to a more friendly role in collection development via taxonomies and classification methods that reinforce how information intelligence is now evolving.

References

[i] Cataloging Policy and Support Office, Library of Congress, *Library of Congress Classification Outline*, http://www.loc.gov/catdir/cpso/lcco/lcco.html
[ii] Pratibha A., Gokhale, "Grey Literature Varieties: Definitional Problems," Third International Conference on Grey Literature, November 13-14, 1997, Luxembourg. *Proceedings* from Amsterdam: GreyNet, 1998: 259-273.
[iii] Julia M. Gelfand, "Academic Libraries and Collection Development Implications for Grey Literature," *Publishing Research Quarterly* 13 #2, 1997: 15-23; Greta Siegel, "Capturing Academic Grey Literature: Starting at Home," *Publishing Research Quarterly* 20 #1 Spring 2004: 62-69; and Heather Lehman and Janet Webster, "Describing Grey Literature Again: A Survey of Collection Policies," *Publishing Research Quarterly* 21 #1 Spring 2005: 64-72.
[iv] ARL Statistics for 1999-2000, reported in *ARL Bi-monthly Report # 219*, December 2001 - see http://www.arl.org/newsltr/219/eresources.htm
[v] David M. Levy and Catherine C. Marshal, "Going Digital: A Look at Assumptions Underlying Digital Libraries," *Communications of the ACM* 38, #4 April 1995: 77-84.
[vi] *Oxford English Dictionary Online* - http://dictionary.oed.com
[vii] Alan Gilchrist, "Thesauri, Taxonomies and Ontologies: An Etymological Note." *Journal of Documentation* 59 #1, 2003: 7-18.
[viii] *Ibid*.
[ix] *Major Categories in the Taxonomy of Educational Objectives* (Bloom 1956) - see http://faculty.washington.edu/krumme/guides/bloom1.html and Benjamin S. Bloom and David R. Krathwohl, *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain.* New York: Longmans, Green, 1956; and David R. Krathwohl, Benjamin S. Bloom and Bertram B. Masia, *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain.* New York: David McKay Co., Inc., 1964.
[x] Joan Dalton, and D. Smith, (1986) "Extending Children's Special Abilities – Strategies for primary classrooms" see also Applying Bloom's Taxonomy at: http://www.teachers.ash.org.au/resarchskills/dalton.htm
[xi] Edward R. Tufte, *Envisioning Information*. Cheshire, CT: Graphics Press, 1990.
[xii] Ben Shneiderman, "The Eyes Have it: A Task by Data Type Taxonomy for Information Visualization," *Proceedings of the IEEE Symposium on Visual Languages*, 1996: 336-343.
[xiii] *Ibid*: 337.
[xiv] *Ibid*: 341.
[xv] Christine Borgman, "What are Digital Libraries: Competing Visions," *Information Processing and Management* 35, 1999: 227-243; Tefko Saracevic and Paul B. Kantor, "Studying the Value of Library and Information Services. : Part II Methodology and Taxonomy," *Journal of the American Society for Information Science* 48 #6 1997: 543-563; Lisa Covi and Rob Kling, "Organizational Dimensions of Effective Digital Library Use: Closed Rational and Open Natural Systems Models," *Journal of the American Society for Information Science* 47 (9): 672-689; Michael Lesk, *Practical Digital Libraries : Books, Bytes, and Bucks.* San Francisco: Morgan Kaufman, 1977.
[xvi] Samuel Demas, Peter McDonald, and Gregory Lawrence, "The Internet and Collection Development: Mainstreaming Selection of Internet Resources," *Library Resources and Technical Services (LRTS)* v. 39 #3 1995: 275-290.
[xvii] *Ibid*: 289.
[xviii] Delphi Group, *Information Intelligence: Content Classification and the Enterprise Taxonomy Practice.* Boston, MA: Delphi Group, 2004:11.
[xix] Bonnie Nardi and Vicki L. O'Day, *Information Ecologies: Using Technology With Heart.* Cambridge, MA: MIT Press, 1999: 49.
[xx] Thomas Davenport with Laurence Prusak, *Information Ecology: Mastering the Information and Knowledge Environment.* NY: Oxford University Press, 1997: 41.
[xxi] *Ibid*: 9.

Additional References consulted include:

Casserly, Mary Frances, "Developing a Concept of Collection for the Digital Age," *portal: Libraries and the Academy* 2 #4, 2002: 577-587.

Gelfand, Julia M., " Challenges for Collections in New Collaborative Teaching and Learning Environments: Does Grey Literature Fill a Void?" Paper Presented at the Seventh International Conference on Grey Literature, Nancy, France, December 5, 2005.

Harloe, Bart and Budd, John, M., "Collection Development and Scholarly Communication in the Era of Electronic Access," *Journal of Academic Librarianship* 20 #2, 1994: 83-87.

Hazen, Dan C., "Collection Development Policies in the Information Age," *College & Research Libraries* 56 #1, 1995: 29-31.

Kovacs, Diane and Elkordy, Angela, "Collection Development in Cyberspace: Building an Electronic Library Collection," *Library Hi Tech,* vol. 18 #4, 2000: 335-361.

LaGuardia, Cheryl and Bently, Stella, "Electronic Databases: Will Old Collection Development Policies Still Work?" *Online* 16 #4, 1992: 60-63.

Norman, O. Gene, "The Impact of Electronic Information Sources on Collection Development: A Survey of Current Practice," *Library Hi-Tech* 15 #1-2, 1997: 123-132.

east view information services

is GREY literature

From Russia

including distant cities and the new independent areas:

Documents issued by government, academia, business, and industry. Low press run and rare books published by Academies of Science, private research groups, business organizations, governmental bodies and institutes. Independent research and the works of advisory bodies. Dissertations and abstracts ("avtoreferats"). **Formerly classified works released through East View's efforts.** Rare and out-of-print from Arkhiv Pechati on microform. Manuscripts, formerly unpublished made available by East View on microform.

From China

and other Eurasian areas

• Social sciences • Economics • History • Law • Philosophy • Literature • Politics • Education • Medicine • Technology • Agriculture

Online, in full text, with English index and abstracts. Search, browse, print and download Conference proceedings, national and international, from academic associations, industrial organizations, governmental agencies, and international bodies.

HARD TO FIND SOURCES IN PRINT, ONLINE AND MICROFORM

Search and browse at **www.eastview.com**

Or, contact East View to set up a highly efficient APPROVAL PLAN

info@eastview.com

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

Nikkia Anderson, Gail Hodge, and Andrea Japzon
Information International Associates, Inc., NASA Goddard Space Flight Center, 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 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.

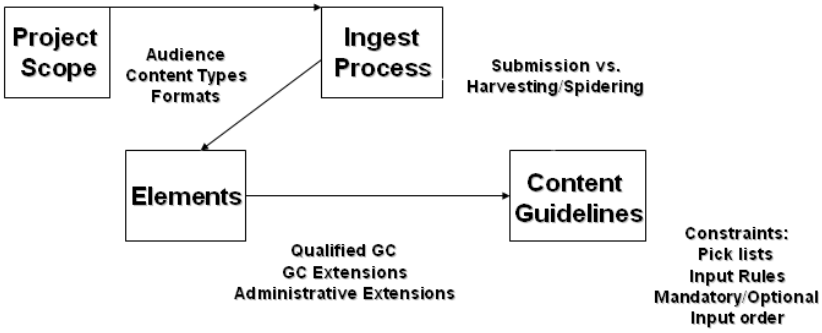


Figure 1: DAS Framework

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]

Hyperactive Grey Objects

Keith G. Jeffery,
CCLRC, United Kingdom

Anne Asserson,
UiB, 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.

This surely is harnessing the power of grey.

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).

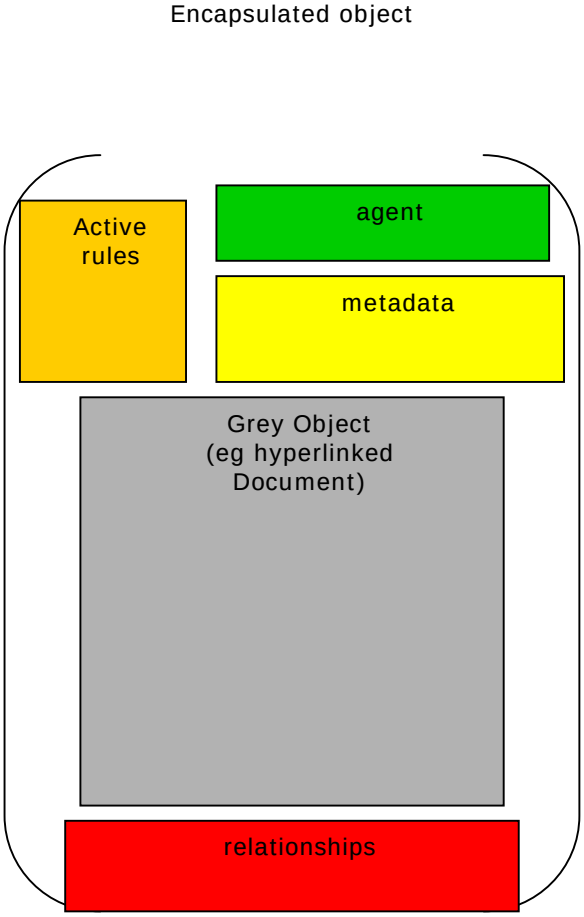


Figure 1

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.

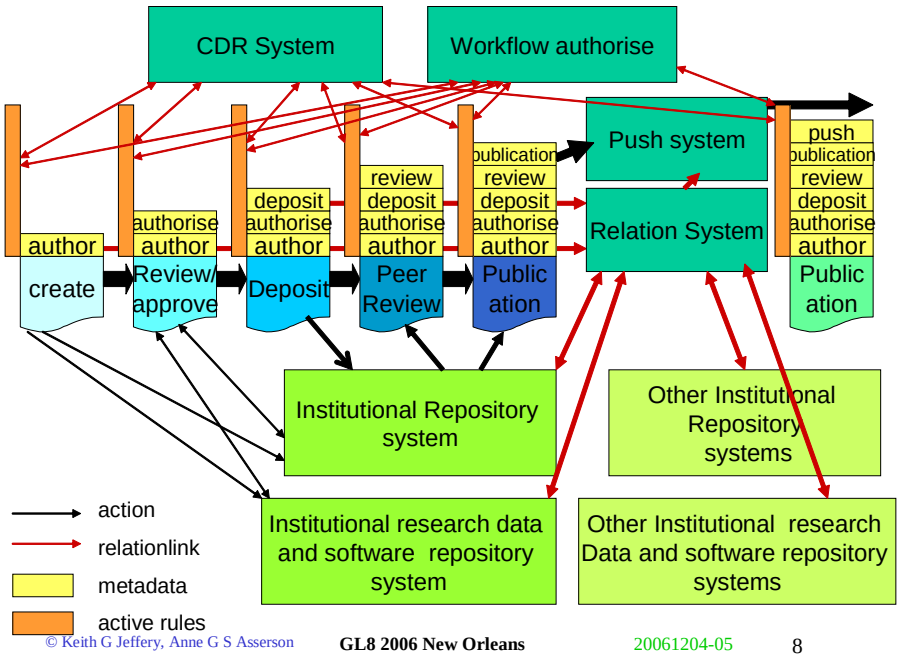


Figure 2

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.

This surely is harnessing the power of grey.

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

Metadata-based analysis to improve clinical trial exchange

Daniela Luzi and Fabrizio L. Ricci,
IRPPS-CNR, National Research Council - Institute of Research on Population and Social Studies, Italy

Luca Dan Serbanati
Politehnica University, Romania

Abstract
There are various, important information sources devoted to the diffusion of clinical trials, but they fail to achieve a complete coverage of clinical research. The demand for a mandatory public registration of clinical trials is emerging from different institutions, which are making efforts to develop common metadata schemas to both increase information exchange and make this information publicly available. The paper describes a metadata analysis of the various solutions of CT data representations adopted by important stakeholders such as national Health authorities, database information providers and standardisation organisations.

1. Introduction

Progress in medical research depends largely on the results of clinical trials (CT). In controlled conditions, and following methodologically sound procedures, CTs test specific clinical and therapeutical hypotheses on a sample of a pre-defined number of patients. CTs are mainly planned and funded by pharmaceutics firms with an interest in testing new drugs or treatments, or by government agencies that back medical research.

Each state has its own regulatory mechanisms for clinical trials along with a corresponding policy for the diffusion of information on CTs. The 1964 Helsinki declaration, successively modified in 2004 [1], established the basic ethical principles that guide medical research when human subjects are involved. Among these principles, the declaration included that of the right to correct information. Regrettably, despite the existence of numerous databases that supply information on CTs, there is no comprehensive source of information on ongoing and concluded clinical studies and even when national registries are maintained, they are often accessible only to a limited number of bodies and regulatory agencies.

The internationalisation of the pharmaceutics industry on the one hand, and of medical research on the other, today raise a series of issues, ranging from the need to disseminate information on clinical trials, so as not to duplicate costly research and to verify the results, to cutting experimentation time in order to increase medical knowledge to benefit public health. The drive to fulfil these objectives involves many stakeholders such as international bodies, medical and patients' associations, the pharmaceutics industry, governmental research bodies and ICT and information specialists. The latter ones need to apply the appropriate technology to support the automation of the complex CT process as well as to develop systems for diffusing information both to potential patients and health professionals.

In the light of this, it is increasingly important to develop interoperable systems based on meta-data that can be easily exchanged among the different components of the CT process, simultaneously enabling diffusion of data on clinical trials also to external users. This explains the urgent demand from numerous stakeholders for CT registration to be made obligatory and for public access to such information. This is not only to fill an information gap, but also to lay the foundations for the development of a common language for information exchange.

This paper presents part of an analysis included in a larger project, which has a twofold aim:

a) the development of a comprehensive CT model which makes it possible to identify suitable tools to automate the entire process;
b) ICT support to increase interoperability between organisations, platforms and applications.

To reach the first goal we have modelled the interaction between the CT sub-processes [2], we have identified the roles and information needs of the stakeholders directly participating in the process (co-ordinating centre, investigators, statisticians, etc.) as well as the information needs of the stakeholders *outside* the process (National Health authorities, systematic reviewers, physicians, patients looking for alternative care treatments).

The second objective, described in this paper, requires a preliminary analysis of the various solutions of CT data representations adopted by important stakeholders such as national Health authorities, information providers and standardisation organisations, which are currently making efforts to develop a common

meta-data schema to increase information exchange and data harvesting from the already existing protocol databases and/or national registries[1].

2. The role of protocol

CTs are defined by the Guidelines for Good Clinical Practice [3] as "any investigation in human subjects intended to discover or verify the clinical, pharmacological and/or pharmacodynamic effects of an investigational product(s) […] with the object of ascertaining its safety and/or efficacy" The approval of a new medical study is based on a highly structured document, the CT protocol, that "describes the objective(s), design, methodology, statistical considerations, and organization of a trial […] and usually also gives the background and rationale of the trial [4]". This grey literature (GL) document has a central role in all the different phases of the complex CT process.

As is clear from this definition, the CT protocol is central throughout the phases of the complex CT process:

- It is the document on which scientific and ethics committees base their approval and as such, is the starting point of any experimentation.
- During the execution phase, it is the guideline document that provides a workplan for all participating centres, detailing all the clinical and diagnostic processes to be carried out on the patient. The workplan must be strictly adhered to so that the CT has comparable results (any deviations must be reported in conformity with pre-established procedures set out in the protocol itself).
- During the evaluation phase, it is the reference document for methodology (definition of patient conditions, selection of evaluation times and data gathering) and for statistical criteria (indicating the type of statistical analysis to be carried out on the data).
- Lastly, the protocol contains information on the various roles of the different organisations participating in the experimentation (names and addresses of sponsors, of any co-ordinating body, of the participating centres, of the laboratories etc.). As such, it is the organisational framework and indicates the structure behind research management.

Consequently, the protocol is the key source of information not only for those participating directly in the research project, but also for the outside stakeholders, such as National Health authorities, systematic reviewers, drug regulatory agencies, researchers and patients. They are all keen to know what experimentation is underway on specific pathologies, on what new investigational products research is being carried out, following what procedure and who is conducting the research and in which hospital. These considerations highlight an important feature of this document, namely that it requires not only a bibliographic description in order to identify it unequivocally (a question that as we will see, raises problems in itself), but that it must be possible to extract specific parts of the document to meet the various users' information needs. Therefore, the analysis of metadata presented in this work includes both a description of the document in the strict sense and a description of the information contained in the body of the document and in some of its chapters.

The CT process closes with the gathering of the results of the clinical data on the participating patients. These are sent via the CRF (Case Report Forms) from the participating centres to the Coordinating Centre or to the Contract research organisation (CRO). The sponsor is responsible for drawing up the clinical study report, which depending on various national regulations, is sent to the reviewing authorities. Guidelines have been set out for this category of document too [5] so as to guarantee correct elaboration of the results. Finally, a regrettably limited proportion of the results of some CTs are published in scientific journals and benefit from a wider diffusion. It has been calculated that a significant percentage of results never see publication at all and that knowledge of both positive and negative results from trials is lost.

In order to provide information on both the entire CT process and the data in the protocol, databases must have access to other sources of information. It is fundamental that they include data on the advancement of the process, such as progress in patients' enrolment, the precise dates of updating for the reported information and finally that they include links to publications or web pages containing CT results. Acquisition of this data is particularly challenging for the database managers given that they must be maintained and updated throughout the entire life cycle of the trial.

[1] In this paper registry and database are used as synonyms. Even if registry mainly denotes that the information provider is a regulatory authority or an agency, there are registries that are functioning as publicly available databases. Moreover, the completeness of information that the word registry should denote is achieved only in some national, not publicly available databases.

3. The role of the registry

The demand for making registration of ongoing and concluded CT protocols mandatory had already arisen by the middle of the '80s [6, 7, 8] and mainly came from the need to correctly assess the results of clinical research. Giving the cumulative acquisition of medical knowledge, it is necessary that systematic reviews, on which changes and developments of the medical practice are based, be able to evaluate the effect of an intervention using a compilation of all relevant findings of all CTs investigating a specific pathology and applying similar criteria and procedures. Some surveys have indicated that this does not always happen [9,10, 11]. It has been shown that roughly half of all trials are never published in scientific journals. There is a tendency for trials illustrating beneficial effects to be published more frequently than those with negative or uncertain outcomes. This leads to overrating the benefits of a specific treatment. Furthermore, research results are published three to five years later and this delay is lengthened in the case of negative or uncertain results [12, 13]. Half of the abstracts presented in medical meetings are never published [14]. This publication bias may and has had negative effects on medical practice and negative consequences for patients, because it can lead to misleading conclusions, overestimating or underestimating certain types of clinical data.

Currently there is a high number of databases that gather CT protocol information. They are managed by regulatory agencies or research institutions, or by pharmaceutical industries and generally contain information on a particular pathology or describe the CT protocol funded by the database information provider. None of these databases provides a comprehensive, international information of CT being currently conducted or already closed.

A comprehensive registry of all ongoing and concluded trials would make it possible to track relevant CTs throughout the multiple existing databases, monitor their advancement, provide important information to access results, whether published or not. Moreover, a public registry can facilitate the retrieval of important information contained in the CT protocol, which can be compared with those written in the clinical study report. This comparison enables a more precise evaluation of CT results, for instance the conformity of initial intents, and of the predefined endpoints of evaluation. More generally, a central registration on trials at inception would also benefit sponsors, regulating bodies and review boards, helping them to avoid unnecessary duplication of CTs and waste of resources.

National regulations for clinical research as well as different policies for the diffusion of health information outline a very disparate situation. Improvements in harmonisation of protocol contents, as well as in clinical report studies have been reached internationally [3, 5]. However, the demand for a global registry is strongly felt [7, 10, 15, 16, 17], but has to face different issues. An international publicly available registry has to be supported by different stakeholders, sometimes with contrasting interests, who have to agree on the type of information to be diffused, the timing of its disclosure as well as on the organisation maintaining and updating the public available registry.
A step towards a national registry took place in 1997 in the US with FDA Modernization Act, Section 113 (FDAMA 113). This act [18] establishes mandatory protocol registration for both federally and privately funded studies conducted under an investigational new drug application (IND) for "serious or life threatening diseases". It also established the development of a publicly available registry, which started to be operational in 2000, the ClinicalTrials.gov [19]. Yet due to lack of enforcement, this registry does not contain all eligible trials [20]. In 2005 the Fair Access to Clinical Trials Act (FACT) was presented to the US Senate to provide for enforcement and to require also registration of device studies in ClinicalTrial.gov. The enforcement of FACT act would also mandate the reporting of all clinical trials results, both positive and negative.

In many European countries (France, Spain, Italy and the Netherlands) it is mandatory to register CT protocols, but such data is only available to national health authorities of drug regulatory agencies.

At a European level a similar procedure has been adopted. In 2001 the European Clinical Trials Directive introduced legislation requiring the registration of "clinical trials on medical products for human use" in a European database [21]. Since 2004 the EudraCT database, supervised by the European Medicines Agency, has been collecting data on all the trials conducted in European member states and gives them a unique identification number. This database is however, accessible only to regulatory agencies and organisations that provide funding for research.

4. Actions supporting mandatory trial registration

In 2004 unreported trial data on the harmful effects of an antidepressant drug used for children was legally pursued in the US for misreporting data trials and this gained widespread public attention [22]. In

truth, among mandatory trial registration advocates harmful effects caused by publication bias were mentioned in various articles in previous years too [10, 11].
2004 was also the year that saw the undertaking of important initiatives, such as that conducted by the World Health Organisation (WHO) as well as statements issued by important stakeholders, who, seeking to exploit their role in the CT process, sought to expedite the process of obligatory disclosure of trial information [23]. It should be noted that each of these stakeholders proposes a set of data that are indispensable for the registration of trials (some of which are analysed in this work) and identify organizations or structures as possible information providers.
Among these stakeholders it is worth mentioning the International Committee of Medical Journal Editors (ICMJE), which includes 11 of the most prestigious medical journals (such as Lancet, New England Journal of Medicine etc.). This Committee made an important statement requiring registration in a publicly available CT registry "as a condition of consideration for publication" [24]. They also establish that the registration should be at or before the onset of patient enrolment. This requirement applied to CTs starting enrolment after July, 1, 2005 and for already ongoing trials the registration due date was set by September 13, 2005. The data set proposed by ICMJE has been successively revised and made compatible with the data set identified by WHO.

If we consider the position of some editors toward past and present initiatives on free open access in other disciplinary fields, this statement is clearly indicative of a remarkable change in the direction of timely and free access to information. Horton [25] has dryly observed that the editor's point of view might change when, as already scheduled by the WHO Platform Regisry, trial results will be diffused in an open access database. Will editors in this case renounce the "Ingelfinger rule" [26] that considers a manuscript for publication only if its substance has not been previously submitted or reported elsewhere?

One of the most influential initiatives is the WHO support and concrete actions for trial information disclosure, for the public good and scientific interests. In 2003, the WHO began a series of consultations with representatives of governments, academia, pharmaceutical industries, patient and consumer groups and medical journal editors that led to a general consensus on the WHO playing a leadership role to in fostering clinical research transparency, as well as coordinating efforts for clinical trial registration. Important statements have been pronounced in successive meetings [27], which can be summarised in terms of the need for global, publicly available, mandatory trial registration through an unambiguous trial identification number and through a set of information both on protocol and trial results made publicly available. In 2005 the International Clinical trials Registration Platform Secretariat was created and became operational pursuing a consensus-based approach. The Who's Registry Platform project aims to develop a search portal "establishing a network of internationally acceptable Primary and Associate Registers, a universal trial reference numbering schema and data interchange standards" Significant outcomes of the project reached at the Geneva meeting in 2005 [23] are the definition of the trials to be registered (interventional trials including early-phase studies) and a 20-item minimum data set required at the time of trial registration and before recruitment of the first patient. Scheduled important agreements include transparency and diffusion on trial information, concerning the timing, format and content of result disclosure, to be met, as above briefly mentioned, not only by pharmaceutical commercial interests, but also by editorial publishing policy [28]. It is evident that the WHO consensus-based approach has to be based on common agreements and also compromises between different needs and frequently contrasting interests.

A "Joint position on the disclosure of sensitive information via clinical trials registers" [29] was established by the innovative pharmaceutical industry, represented by the European Federation of Pharmaceutical Industries and Associations (EFPIA), the International Federation of Pharmaceutical Manufacturers and Associations (IFPIA), the Japanese Pharmaceutical Manufactures Association (JPMA) and the Pharmaceutical Research and Manufacturers of America (PhRMA). This document sets out the Association's commitment to registering all information about all new ongoing trials in a free, publicly accessible clinical trial register as well as the intention to "post results on a drug that is approved for marketing and is commercially available". Even if the association accepts in principle the WHO minimum data set, it expresses concerns on some data fields "regarded as sensitive for competitive reasons by sponsors who may wish to delay the release of the information. (The fields mentioned are; primary outcome; key secondary outcomes; intervention; target sample size; and official scientific title). Moreover, in excluding the "exploratory trials" from the mandatory registration, the association proposes that sponsors may provide copies of protocols and amendments to medical journals, upon request, entering "into confidentiality agreements" with them to ensure protection of information in protocols and amendments.

A more radical intervention for "ensuring transparency to fulfil ethical obligations and minimal bias" was expressed in the Ottawa Statement endorsed by more than 100 people and organisations world-wide. The draft statement was conceived in 2004 during the 12th Cochrane Colloquium. The statement grounds the mandatory trial registration on a list of ethical and scientific principles [30, 31]. It sustains the registration

of all types of trials "related to health or healthcare regardless of topic, design, outcomes or market status of intervention examined" as well as a unique identification number. It also identifies the time lines during which information on the protocol, its amendments, process status and results should be registered and asserts that results should be made publicly available promptly, regardless of their publication status. Although the Ottawa group supports the efforts of WHO and JCMJE as well as the 20 WHO minimum data set, it requires that a more detailed list of information related to the protocol as well as trial results should be made publicly available. It also requires making the full text of the protocol, including its amendments, publicly available.

5. Method

The analysis of meta-data representing both the protocol and the CT process took as its first step the identification of data schemas developed by stakeholders, which had to represent different environments. From the comparison of the data schemas it was evident that there were too many differences in headings as well as in groups of data sets. For this reason we developed our own *reference schema*, which could encompass all the metadata contained in the stakeholder schemas.
The analysis of the metadata has considered the following stakeholders:

- National and international regulatory bodies that have proposed or constructed CT management systems;
- Bodies working on data standardisation in order to facilitate information exchange and to promote interoperability between organisations, platforms and programmes;
- Organisations granting access to CT databases. Out of the various existing databases we have selected those from different geographical areas or those with a lengthy experience in the gathering and distribution of CT data on specific pathologies.

The metadata analysis has taken into consideration:

- The name of the data element, Its description or semantic definition of the metadata, The values or extensions of the data elements. Then we mapped the data elements with stakeholders' information needs.

5.1. Stakeholders providing data schemas

Our analysis takes into consideration data schemas developed by:

- International regulatory bodies,
- Organisations promoting standards, and
- Public available protocol databases.

Below, we give a brief description of each organisation.

A) International Regulatory authorities

- International Conference on Harmonisation of Technical Requirements for registration of Pharmaceutical for Human (ICH) was established in 1990, on the initiative of the then European Community. ICH's commitment is "to increase international harmonisation, aimed at ensuring that good quality, safe and effective medicines are developed and registered in the most efficient and cost-effective manner". The harmonisation is achieved within Europe and through bilateral contacts between Europe, Japan, the USA and other regions. Among the guidelines developed within ICH, the Guidelines for Good Medical Practice ICH E6 [3] is the international reference point for the compilation of a protocol, while the Guidelines for Good Medical Practice ICH E3 [5] is used to deliver trial results. We have analysed the ICH E6, although it is not a genuine database schema, because of its importance, for the completeness of the information description, which has to be contained in a protocol as well as for its internationally accepted definitions.

- World Health Organisation - International Clinical Trials Registry Platform (ICTRP), as above mentioned, was established in 2005, on the basis of the Ministerial Summit on Health Research, which took place in Mexico City in 2004 and on the recommendations presented at the 115th WHO executive Board. It is a WHO project which involves the Research Policy & Cooperation Department (RPC) in the Evidence & Information for Policy Cluster (EIP). The Registry Platform is overseen by an International Advisory Board and by a Scientific Advisory Board. The Registry Platform aims to set "international norms and standards for trial registration and reporting that uphold scientific and ethical good". The 20 data set [32] analysed in this paper constitutes a fundamental reference point for the harmonisation of existing and future CT protocol databases, whose compliance could greatly improve data sharing and interoperability.

- European Clinical Trails database (EudraCT) The above mentioned European Directives 2001/20/EC established the development of the EudraCT database, which became operational in 2004. The database is confidential and is only accessible by Competent Authorities of the European Member states, The EMEA and the Commission. A "Detailed guidance on the European clinical trial database" [33] reports user requirements, data submission and communication procedures, along with search functions. We have focused our analysis on the core dataset as well as on its description given in this document, comparing it to the application form [34] used to submit trial information and to obtain the EudraCT protocol number. This data set was also used as a reference point for our analysis on stakeholders' information needs.

B) Organisations promoting standards

- Among the various organisations working on medical standards we have decided to analyse the Clinical Data Interchange Standards Consortium (CDISC) for several reasons. CDISC is a non-profit organisation, which has set up different international groups working on the development of data model in various aspects of the entire CT life cycle. The approval of each data model is subject to a ballot procedure, which gathers comments from other teams. The consortium is "committed to the development of worldwide industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trials data and metadata for medical and biopharmaceutical product development. CDISC is also actively co-operating with other organisation promoting standards, such as HL7, which is committed to the standardisation of clinical data messages and with caBIG™ (Biomedical Informatics Grid™), an open-source, open-access information network enabling cancer researchers to share tools, data, applications, and technologies. At the moment CDISC has developed an operational data model, a laboratory model, and a study data model for tabulation. Its protocol representation contains 264 data elements [35], with definitions taken from a glossary, when available, as well as descriptions and values. Its comprehensive approach to the CT can help both data sharing between registries and automated support of the entire process.

C) Public available protocol databases

- ClinicalTrials.gov became operative in 2000 as a result of the 1997 FDA Modernization Act Section 113 (FDAMA 113). The system is managed by the U.S. National Institutes of Health (NIH), through its National Library of Medicine (NLM), which revises the submitted information and updates the database daily. As required by the law, the database hosts federally and privately funded studies under an investigational new drug application (IND) for "serious or life threatening diseases". As there is no clear definition of "serious or life threatening diseases", the universe of trials that should be registered cannot be identified. Moreover, the law does not provide a specific enforcement mechanism [20], so the registry is not so comprehensive as it should be. Since 2004 the registry has allowed voluntary registration from other countries on the condition that protocols have been approved by a review board and conform to the regulations of the appropriate national or international health authority.
  Data is submitted by the CT sponsors via a web-based protocol registration form [36]. Submissions are evaluated by NML staff, and key information is revised using Medical Subject Headings (MeSH) controlled vocabulary. Moreover, the publicly available web site provides users with additional information and links to both health and non-health professional sources of information.

- Clinical Trials PDQ is managed by the US National Cancer Institute (NCI), which coordinates the National Cancer Program and funds either intramurally and extramurally CTs on cancer treatment, prevention and detection. The database [37] includes most of the NCI funded trials, but there is no mandatory rule for investigators to submit trial information. The voluntary protocol information registration is assured by an online submission format. Protocol information is publicly available in two formats, the health professional version, which includes more detailed information in a technical language, and a patient version which is simplified both in language and amount of information. PDQ regularly exchanges its data with ClinicalTrials.gov and vice versa. In the analysis of the meta-data required by PDQ [38], we have analysed the health professional version, while we have compared the patient version with the results obtained in our analysis on stakeholders' information needs.

1) Current controlled clinical trial (CCT) Ltd [39] is part of the Current Science Group of biomedical publishing companies, which includes BioMed Central. In 1998, CCT launched a web site providing free access to Controlled trials, i.e. studies where the investigational product is compared with a treatment that has known effects or with a placebo. Since 2003 CCT has allowed registration of CT protocols and assigns a unique 8-digit number, the International Standard Randomised Controlled Trials Number (ISRCTN) and makes this information available in the *metaRegister*. This meta-register contains information from UK Medical Research Council, the NHS Trusts Clinical Trials Register, the US National Institutes of Health (NIH) and the NHS Research and Development Health Technology Assessment Programme (HTA) and from other organizations which voluntarily submit information. The data

schema [40] analyzed in this paper is the one developed to provide the ISRCTN, and connected to the metaRegister.

- Dec-Net register [41] is a relatively new register which involves trials from 4 European countries (France, Italy, Spain, UK) and has been developed with the support of the Fifth Framework Programme of the European Community [42]. It aims at promoting communication and collaboration among researchers, disseminating CT results and facilitating patient access and recruitment to trials. It has been chosen for the meta-data analysis, because is one of the few, if not the only one, devoted to trials involving children. Moreover being a co-operative database it might have already taken common and exchangeable data elements into consideration.

6. Results

The analysis of the data schema produced by the chosen stakeholders did not enable us to select one particular schema, because they had:

- different headings under which a set of data elements were grouped, and
- each group of data sets had different types of information

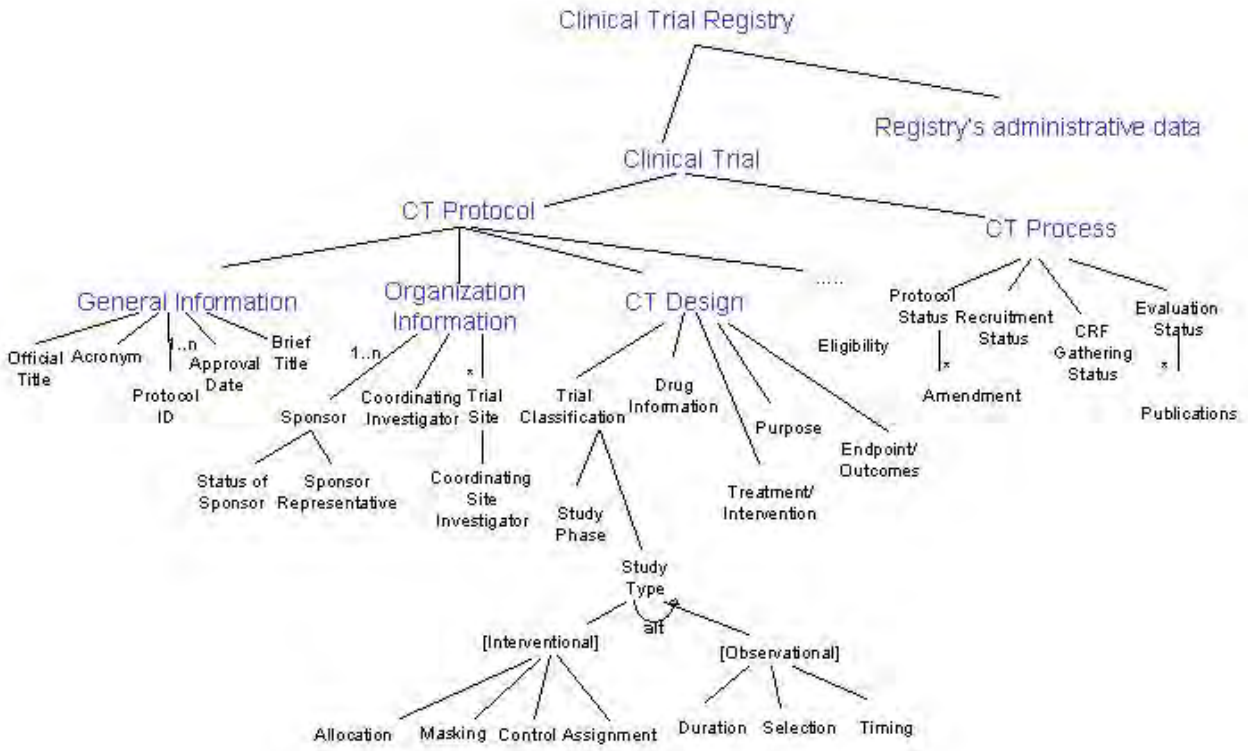


Fig. 1. - Information Structuring schema of CT registriies

For this reasons we had to define our data model (fig. 1), which had to be able to contain all information types present in the union of each data schema. In choosing a name and its definition we have selected those most commonly used. In this way our model represents a reference point which enables the comparison among the stakeholders' meta-data elements. This reference schema takes only a high level of data elements into consideration. A more detailed analysis of the meta-data would have shown more differences both in data names and descriptions, as we noted when we tried to analyse in particular data elements such as drugs or treatments. We have highlighted this difference using VDI in the table (see table description below).

Moreover, the schema does not include protocol information such as statistic methods, assessment of efficacy and safety, which are never reported in public protocol databases, even if they contain important information for the evaluation of CT results.

The tables describe two different types of information:

- Presence (described by the sign "√") or absence of a meta-data (described by the sign "--"). We have also reported additional information such as "not publicly available" (NPA for short), or optional information (O. for schort). The sign "√" also represents homogeneity with the data element chosen in the reference schema.
- The meta-data comparison considers the name of the meta-data, its description and the values or extensions associated to it. To represent this comparison we used the following acronyms:
    - DN = different names are used for the identification of the same data element;
    - DD = there is the same name of the meta-data, but a different description of the data element;
    - DE = the values describing the data elements belong to different definition sets;
    - DC = different classification and consequently different codes are used to describe the same data element.

We also used the expression "very detailed information" (VDI) to underline that a data element is described in a very detailed way, generally with a certain number of additional values.

Table 1. - Protocol General Information

General information	ICH E6	WHO	EudraCT	CDISC	Clinical Trials.gov	PDQ	ISRCTN	DECNet
1.Protocol title	√	DN	√	√	DN	DN	√	√
2. Acronym	--	√	DN	--	--	--	√	√
3. Protocol IDs	DD	VDI	DD	VDI	DD	VDI	VDI	VDI
4. Approval date	DD	--	DN	√	--	--	--	--
5. Brief Title	--	DN	--	--	√	ND	--	--

Table 1 shows data set describing the general information about CT protocol. The main problem is represented by the protocol identification number (ID). Generally this code is directly assigned by the sponsor and the definition of "primary ID" should identify it. However, many IDs may be assigned to protocols. For instance in Europe the mandatory registration produces the assignment of an EudraCT ID number and the voluntary subscription to the ISRCTN provides another ID code. Each number is generally preceded by the acronym of the organisation providing the ID, nevertheless, the electronic management of these different codes is not easy and does not always help to identify the protocol unambiguously. Many protocol databases provide all the protocol IDs, giving very detailed information (VDI).
In addition, the protocol title may be identified with different names (DN) by the various databases. The databases oriented to provide protocol information to health professionals call it "official", "scientific" or "full title" to differentiate this title from that elaborated to give information to non-health professionals. This latter title is reported under different heading such as "lay title" or "public title". The first one generally has a *dense* information content, reporting study phase, pathology name, drug and comparative products and/or treatments as well as health conditions of eligible patients. Each item of this information could be classified by means of standardised tags, such as XML, already at the beginning of the study, but this has the "disadvantage" of disclosing sensible information such as investigational products.

Table 2. – Information on the CT organisation

Information on organisation	ICHE6	WHO	EudraCT	CDISC	Clinical Trials.gov	PDQ	ISRCTN	DECNet
1.Sponsor	√	√	√	√	√	√	√	√
1.1 Sponsor status	--	DN	√	√	--	--	--	√
1.2.Sponsor representative	√	--	√	√	--	--	--	--
2.Coordinating investigator	√	DD	--	√	DE		--	√
3. Trial sites	VDI	DN, DE	DD	√	VDI	VDI	DE	VDI
3.1.Co-ordinating site investigator	√	--	√	√	--	DD	--	--
Contact person	--	√	--	--	√	√	--	--

Trial planning and execution are complex activities carried out by many stakeholders, each one with a precise responsibility and role. The meta-data set reported in table 2 gives only a restricted set of information provided in a protocol. Sponsor information is always given in the examined databases, some of which specify if the sponsor is a commercial or non-commercial organisation. Especially in the public databases willing to provide information to prospective patients, there is very detailed information in the data set pertaining to trial sites information (organisation names, full addresses, web sites of the CT participating centres). Sometimes the values of the meta-data trial sites may vary providing only the name of countries involved in the study or the number of CT participating countries (DE).

Table 3. – Trial Design

CT Design	ICHE6	WHO	EudraCT	CDISC	Clinical Trials.gov	PDQ	ISRCTN	DECNet
1 Study phase	--	--	√	√	√	√	--	√
2. Study type	√	DE	√	√	√	√	--	--
2.1. Interventional	√	--	√	√	√	--		--
2.1.1. Allocation	√	√	√	√	√	DD	√	--
2.1.2. Masking	√	--	√	√	√	--	√	√
2.1.3. Control	√	√	√	√	√	--	√	--
2.1.4. Assignment	√	--	√	√	√	--	√	--
2.2. Observational	--	--	--	√	√	--	√	--
2.2.1 Duration	--	--	--	√	√	--	√	--
2.2.2. Selection	--	--	--	√	√	--	√	--
2.2.3. Timing	--	--	--	√	√	--	√	--
3. Drug information	√	DN, DE	VDI	VDI	NPA	--	--	VDI
4. Health conditions	DN	√	DC	√	DC	VDI	DE	DC
5. Treatment/ interventions	DE	DE	--	VDI	DE	√	DE	--
6. Objectives	√	--	√	VDI	DE	√	DN, DE	ND
7. Endpoints/outcomes	DN, DE	DN, DE	√	VDI	√	√	√	√

Table 3 contains the kernel of protocol information. Especially for health care professionals, this information gives important details on what the trial is going to evacuate, (for instance the effectiveness in phase II trials), and how patients will be assigned to intervention groups (such as randomisation, blinding procedures described by the meta-data 2.1 - 2.2.3). Table 3 shows that this type of information is generally available in regulatory registries and in CDISC, while among the publicly available databases, only ClinicalTrials.gov and ISRCTN provide this information. However, it should be noted that ISRCTN requires this information to be reported in a free text field and this do not give any guarantee as to how the information will be filled in by the protocol submitters. This also makes the identification and retrieval of information sub-set difficult.
Data on drugs is sensitive information especially in the case of new investigational products. They could be very detailed, as required by regulatory organisations, where the sponsor has to describe chemical drug principles, products characteristics, or, if the product has been already authorised, trade name and code. ClinicalTrials.gov, representing the US mandatory registry, requires this detailed information for the same reasons, but it does not make it publicly available (NPA). The 20 items proposed by the WHO insert this information in the meta-data "intervention(s)" usually used in other schema to describe patient treatments. Moreover the WHO data set requires only the International Non-proprietary Name of pharmaceutical substances.

Concerning Health conditions, the information of the pathology is generally available, but different codes are used, for instance DecNet adopts the ICD-9, EudraCT the MedDRA classification and ClinicalTrials.gov the MeSh thesaurus.

The field "Treatment/interventions" should include the description of treatment schedules and dosage of the investigational products as well a o the as well as of other drugs used in the trial. This type of information is not present in the majority of publicly available databases. In the example of ClinicalTrials.gov the extension of this metadata is different as it requires to select the type and name of intervention from a pre-defined list, without mentioning any other information on dosage or treatment.

Also the information about Endpoints is not always comparable. Endpoints are the pre-defined times and procedures to be carried out to make ad interim and final result analysis. Only the comparison of this information with that contained in the clinical study report can make sure that the final results are coherent with the study hypothesis. Some organisations (EudraCT, ISRCTN) require free text description on primary outcomes and this does not give any guarantee on how the information will be filled in by the protocol submitter.

Table 4. – Trial Eligibility criteria

Eligibility criteria	ICHE6	WHO	EudraCT	CDISC	Clinical Trials.gov	PDQ	ISRCTN	DECNet
1. Inclusion	√	√	√	VDI	√	VDI	√	√
1.1 Age	--	--	√	√	√	√	--	√
1.2. Gender	--	--	√	√	√	√	--	√
2 Exclusion	√	√	√	VDI	--	--	√	√
3 Target size number	√	√	√	VDI	O.	√	DE	√

Table 4 shows a set of data, which is very important for the majority of CT stakeholders. It is very homogenous especially in the case of inclusion criteria, while exclusion criteria are sometimes not reported at all.

Table 5. – Trial process status

CT process status	ICHE6	WHO	EudraCT	CDISC	Clinical Trials.gov	PDQ	ISRCTN	DECNet
1. Protocol writing	√		√	√	--	--	--	--
1.2 Amendments	√	--	DE	√	--	--	--	--
2. Recruitment	--	DE	--	--	DE	DE	--	VDI
3. Summery of results	--	--	--	--	--	--	--	√
4. References to Pub.	--	--	--	--	√	√	√	√
5. Links to websites	--	--	--	--	√	--	√	√

The meta-data set shown in table 5 provides information used to monitor the process. One of the most important items is related to recruitment, which describes if the protocol is closed or still recruiting patients. Most of the databases report this information, but each one uses its own set of predefined values to present such information. References to published literature, which generally contains results of the trial, is also reported only in publicly available databases.

7. Stakeholders' information needs: two examples

The analysis of stakeholders' information needs is based on our previous CT process modelling [2]. In particular the description of the activities that compose the entire CT process has defined the identification of data necessary to carry out each activity as well as the information flow among the different stakeholders participating in the CT process. The description also comprises *use case analysis* in UML (Unified Modelling Language), and this has given us further insight in the development of stakeholders' information needs.

In this paper we have compared the results of our previous analysis with the data elements contained in the CT database. The results are reported in fig. 2 and fig. 3. The patient's view reported in fig 2, considers patients looking for new therapeutic treatments and seeking to participate in CTs. Prospective patients generally would like to know if there are clinical studies in certain pathologies, and if their health conditions are included in the protocol eligibility criteria. In the publicly available databases the name of pathology is one of the search values used to retrieve information on CTs, and the use of MeSh in some databases, such as Clinical.trials.gov, help the prospective patient to retrieve both medical and common language terms. Inclusion criteria are usually reported, even if the language used presumes medical knowledge. Other important information is related to names and addresses of the centres where the trial is conducted (sponsor representative and coordinating site investigator) and contact person, from whom the patient may ask for further information. Information that rarely appears in patient version is that related to objectives, treatment/interventions planned and trial results. In all these cases an effort in reporting information comprehensible to non-health professionals is needed and this implies a revision and "translation" process undertaken by specialised personnel, maybe a librarian review board. Single items of the protocol should in this case be extracted and a standardised protocol meta-data schema could also help significantly when "translating" patient content.

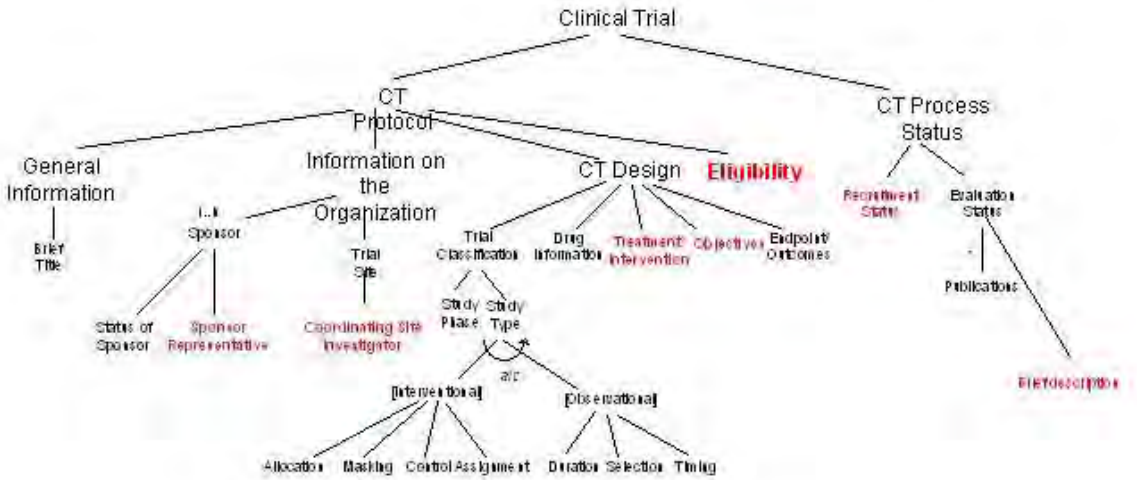


Fig. 2. - Patient's information view

Fig. 3 gives the information view of a National Health Authority that should know which studies are being conducted, on which investigational product, who is carrying out the research and what the results are. In this case, a single protocol ID is essential for tracking all ongoing and closed trials helping the National Health Authority decide on studies to be funded, as well as on the safeguard of public health. Other information needs pertain to very detailed information contained in the protocol, such as drug information and endpoints. The EudraCT meta-data schema analysed in this paper seems to fulfil these requirements, especially in the case of drug information. The request of other information in free full texts may be too vaguely reported and may not correspond to the National Health Authority duties.

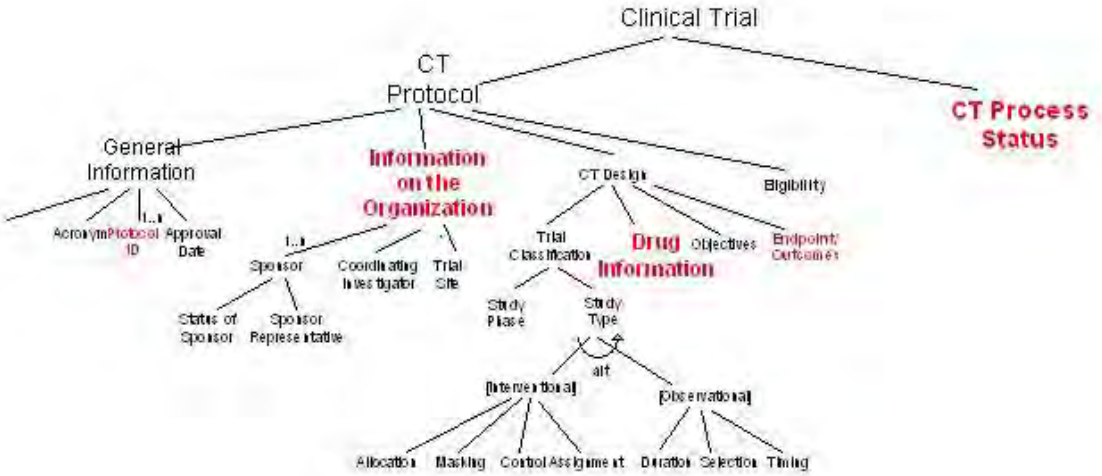


Fig. 3. - National Health Authority's Information view

8. Conclusions

The paper has presented an analysis of meta-data contained in different CT protocol databases. Generally speaking we can say that the type of information source influences the acquisition models as well as the content of the databases. Registries maintained by regulatory authorities, generally not publicly available, contain detailed information on drugs as well as on the organisation and responsibilities of the studies, while public available CT databases tend to focus on information for patients and care providers. CDISC is concerned with a very detailed description of protocol data elements, but does not consider information on the CT process.

The protocol turns out to be the main information source. Its standardisation could lead to a simpler and more uniform data extraction, which could be transformed into a predefined template, making the publicly available information comparable and reliable. To this end a unique protocol ID represents an essential goal, confirming the importance of the actions taken to support the mandatory CT registration.

The analysis of meta-data has highlighted that, although many efforts have been made by different organisations and associations, there are still many differences in the descriptions of the protocol. They are related to the name of the metadata, but also to its definition and extensions. Critical points are information on drugs, as well as endpoints and outcomes.

References

[1] World Medical Association (2004), Declaration of Helsinki. Ethical principles for medical research involving human subjects. Available at: http://www.wma.net/e/policy/b3.htm (accessed 20 November 2006)
[2] Collada Ali, L., Fazi, P., Luzi, Daniela, Ricci, F.L., Serbanati, L., Vignetti. M., (2004) Toward a model of clinical trials, In: Biomedical and Medical data analysis, 5th International symposium, ISBMDA 2004, Barcelona, Spain, B. J., Martin-Sanchez, F., Maojo, V., Sanz, F. (Eds.), (LNCS 3337), Berlin, Springer, 2004.
[3] European Medicines Agency, ICH Topic E 6 (R1) (2002). Guideline for Good Clinical Practice. Note for guidance on good medical practice, (CPMP/ICH/135/95), July 2002, Emea, 59 p.
[4] Clinical Research Glossary (version 4.0), (2005), In: Applied Clinical Trials, 14 (12), 2005, pp. 24-58.
[5] European Medicines Agency, ICH Topic E 3 (1996). Structure and Content of Clinical Study Reports. Note for guiding on structure and content of clinical study reports, (CPMP/ICH/137/95), July 1996, Emea, 48 p.
[6] Simes, RJ. (1986) Publication bias: the case for an international registry of clinical trials, In: J Clin Oncol, 6 (1) Oct 1986, pp. 1529-41.
[7] Tonks, A. (1999), Registering clinical trials, In: BMJ, 319, 1999, pp. 1565-1568.
[8] McCray A.T. (2000), Better access to information about clinical trials; In: Annals of Internal Medicine, 133 (8)17 Oct. 2000, pp. 609-614.
[9] Chalmers, I. (1999), Underreporting research is scientific misconduct, In: JAMA, 263 1999, pp. 1405-1408
[10] Dickersin, K., Rennie, D. (2003), Registering Clinical Trials, In: JAMA, 209 (4), July 2003 pp. 516-523.
[11] Manheimer, E., Anderson, D. (2002), Survey of public information about ongoing clinical trials funded by industry: evaluation of completeness and accessibility. In: BMJ, 325, 2002, pp. 528-531.
[12] Lexchin J., Bero A. L., Djulbegovic B., Clark O. (2006), Pharmaceutical industry sponsorship and research outcome and quality: systematic review, In: BMJ, 326, 31 May 2003, pp. 1167-1170.
[13] Stern J. M, Simes J.R. (1997), Publication bias: evidence of delayed publication in a cohort study of clinical research projects. In: BMJ. 315, 1997, pp. 640-645.
[14] Krzyzanowska M. K., Pintilie M., Tannock, I.F. (2003), Factors associated with failure to publish large randomized trials presented at an oncology meeting. In: JAMA, 290 (4) July 2003, pp. 495-501.
[15] Horton, R.; Smith R. (1999), Time to register randomised trials. In: BMJ, 319, Oct. 1999, pp. 865-866.
[16] Tonks A. (2002), A clinical trials register for Europe, In: BMJ, 325 December 2002, pp. 1314-1315.
[17] Haug C., Gøtzsche P., Schroeder T.V. (2005), Registries and registration of clinical trials. In: New England Journal of Medicine, 353 (26) Dec 2005, pp. 2811-2812.
[18] U.S. Food and Drug Administration (1997), FDA Modernization Act of 1997, available at http://www.fda.gov/oc/fdama/default.htm (accessed 20 November 2006).
[19] Gillen, J.E., Tse, T., Ide, N.C., McCray, A. (2004) Design, implementation and management of a web-based data entry system for ClinicalTrials.gov. In: MEDINFO 2004, M. Fieschi et al. (Eds.), Amsterdam, IOS Press, pp. 1466-1470.
[20] Zarin, D., Bergeris A., Ide N.C., Tse T. (2005), ClinicalTrials.gov: a report to the Board of Scientific Counsellors, *Technical report LHNCBC-TR-2005-003*, 28 p.
[21] European Parliament (2001), Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, relations and administrative provision of the Member States relating to the implementation of good clinical practice in the conduct of clinical trial on medical products for human use. In: Official Journal of the European Communities, L121/34.
[22] Whittington C.J., Kendall T., Fonagy P., et al. (2004), Selective serotonin reuptake inhibitors in childhood depression: Systematic review of published versus unpublished data. In: Lancet, 363 2004, pp. 1341-1345.
[23] Evans, T., Gülmezoglu M., Pang T. (2004), Registering clinical trials: an essential role for WHO. In: Lancet, 363, 2004, pp. 1413-1414.
[24] De Angelis C., Drazen J.M., Frizelle, F.A. (2004), Clinical trial registration: a statement from the International Committee of Medical Journal Editors. In: N Eng J Med, 351 2004 pp. 1250-1251.
[25] Horton R. (2006), Trials registers: protecting patients, advancing trust. In: Lancet, 367 2006, pp. 1633-1635.
[26] Toy J. (2002), The Ingelfinger rule: Franz Ingelfinger at the New England Journal of Medicine 1967-77. In: Science Editor, 25 (6) 2003, pp. 195-196.
[27] World Health Organization, About ICTRP. Available at http.//www.who.int/ictrp/about/details/en/print.html (accessed 20 November 2006)..
[28]World Health Organization, Department of Research Policy and Cooperation World Health Organization (2006), International Clinical Trials Registry Platform, 2nd Scientific Advisory Group Meeting, Geneva, 27-28 April 2006, Meeting Report, 21 p.
[29] International Federation of Pharmaceutical Manufactures and Associations (2005), Joint position on the disclosure of clinical trial information via clinical trials registries and databases. Available at: http://www.sanofi-aventis.com/images/050106_release_en.pdf (accessed 20 November 2006).

[30] Krleza-Jeric, K., Chan, A.W., Dickersin, K., Sim I., Grimshaw, J., Gluud C. (2005), Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). In: BMJ, 330 2005, pp. 956-958.
[31] Ottawa Statement on Trial Registration (2005), Draft – The Ottawa Statements, Part two: principles of operationalisation for international trial registration. Available at http://ottawagourp.ohri.ca/statement2.html (accessed 20 November 2006).
[32] World Health Organization (2006), Trial Registration data Set. Available at: http://www.who.int/ictrp/data_set/en/print.html (accessed 20 November 2006).
[33] European Commission (2004), Detailed guidance on the European clinical trials database (EUDRACT Database), Brussels, ENTR/CT 5.
[34] CT Application Form. Available at: http://eudract.emea.eu.int/document.html#guidance (accessed 20 November 2006).
[35] HL7-CDISC. PR Group (2005), Spreadsheet V2.0. Available at http://www.cdisc.org/standards/ (accessed 20 November 2006).
[36] Data element definition. Available at http://prsinfo.clinicaltrials.gov/definitions.html (accessed 20 November 2006).
[37] http://www.cancer.gov/clinicaltrials (accessed 20 November 2006).
[38] Online PDQ Protocol Submission Instructions available at http://www.cancer.gov/cancertopics/pdq/protocol-submission-instructions (accessed 20 November 2006).
[39] http://www.controlled-trials.com (accessed 20 November 2006).
[40] Current Controlled Trials ISRCTN Guide, ISRCTN Field Definition, available at: http://www.controlled-trials.com/isrctn/applications/fieldDefintions.htm (accessed 20 November 2006).
[41] http://www.dec-net.org/ (accessed 20 November 2006).
[42] Santoro E., Rossi V., Pandolfi C., Bonati M. (2006), DEC-net: the development of the European register of clinical trials on medicines for children. In: Clinical trials, 3 2006 pp. 366-375.

Implications of copyright evolution for the future of scholarly communication and grey literature

Marcus A. Banks
Frederick L. Ehrman Medical Library, United States

Cees de Blaaij
Zeeuwse Bibliotheek, The Netherlands

Introduction

Traditional practices regarding copyright are undergoing transformation. Although it is still common for scholars to give up their rights to their articles so that they will be published, this happens less frequently than it once did. Our analysis of the RoMEO database [1] shows that 75% of publishers allow authors to post their work in an online repository, whether that repository is hosted by their institution or on a personal web page. Whatever becomes of the open access movement to make all peer-reviewed journal articles immediately available online, copyright liberalization represents an enduring legacy of the open access movement.

Online repositories are a more natural home for grey literature than open access journals. Repositories can store working papers and technical reports (among other content types) just as easily as peer-reviewed articles. Crucially, repositories can also store raw data, the *grey content* that lies at the root of much scholarly discovery. Copyright liberalization has encouraged the proliferation of such repositories; one prominent example is arXiv, which primarily serves physicists and computer scientists [2]. As scholarly discourse evolves, the preservation and promotion of grey content should command more energy than providing access to discrete grey literature.

I: Open Access, Self-Archiving and Institutional Repositories, and Open Data

Open Access

An open access publication is freely available to anyone with an Internet connection, and digitally archived to ensure permanent access [3]. The debate about whether to provide open access, and how, continued to evolve in 2006.

Professional societies generally support the goal of open access, which is to maximize the dissemination of scholarly knowledge. By now, the increased exposure that results from open access is empirically indisputable [4]. Despite this clear benefit, many society publishers continue to view open access publishing with ambivalence. Most societies depend on traditional subscription revenues to fund other activities, such as annual meetings. Without a comprehensive plan to replace the subscription revenues that are lost under an open access model, societies have been reluctant to embrace it. Several open access advocates have advanced proposals for how societies can surmount this challenge [5, 6].

At one time commercial publishers ridiculed proponents of open access publishing as starry-eyed idealists who did not know much about the economics of scholarly publishing [7]. Those days are gone. In 2006 several leading commercial publishers (along with society and university publisher counterparts) began to offer a “hybrid” open access publishing option [8].

It is now possible to find open-access articles alongside traditional articles in the same electronic issue of a journal. The open access articles are available to everyone, while the traditional articles require a subscription for immediate access. The authors of each article make this decision themselves. Any fees associated with open access are absorbed by funding agencies, are waived, and are sometimes (not always) paid by the authors [9]. The hybrid model allows savvy publishers to generate several funding streams, while the traditional subscription-based model of paying for journal publication slowly contracts.

Depending upon policy developments around the globe, hybrid open access may yield to complete open access in many cases. In the United States Senate, the “Federal Public Research Act of 2006” seeks to ensure that all articles that result from research funded by the federal government, “in whole or in part,” are available for free online no later than six months after publication [10]. The bill has not passed, as of the time of this writing. It has a great deal of momentum, however, and passage in some form seems likely [11]. This is a strikingly different from the political realities in 2003, when a bill with similar aims—the “Public Access to Science Act”—was quietly buried. In the intervening years, the open access movement has matured.

The European Commission is also taking steps to endorse open access. In a wide-ranging report published in January 2006, the Commission recommends that European funding agencies “guarantee public access

to publicly-funded research results shortly after publication" [12]. The Commission provides numerous practical suggestions for how to do this. A second report, anticipated for December 2006, will expand upon this theme.

The open access movement has increased access to white literature. Nevertheless, it is important for scholars of grey literature to remain abreast of developments in this area. The Internet age is slowly eroding the traditional distinction between grey and white literature. This distinction is ultimately arbitrary, and is fading away.

In the meantime, the self-archiving and institutional repository movements bear more directly upon efforts to increase exposure to grey literature and grey content.

Self-Archiving and Institutional Repositories
The large majority of publishers allow authors to post versions of their articles on their own web site, which is known as self-archiving. Because a self-archived article is not a formally published work, it is a type of grey literature (even if the archived material is very similar to the official publication.) Although many publishers have permitted self-archiving for years, most scholars do not archive their works [13]. Institutional repositories relieve scholars of this archival responsibility, and are designed to preserve more than standard articles. For this reason, institutional repositories have great potential for increasing access to grey literature [14]. But it is not yet customary for researchers to deposit their scholarly materials (including grey materials) into institutional repositories.

Open Data
Widespread adoption of the principles of the "open data" movement should lead to increased use of institutional repositories. The success of the open data movement will be a critical factor in shifting the focus from grey literature to grey content.

The open data movement is a corollary of the open access movement. Just as scholarly articles should receive the widest possible exposure, the data that underlies research results should also be freely available. Although the open access movement has made impressive gains, it is essentially concerned with access to the same type of information that was available in the print-only era. The open data movement is only possible in an electronic environment.

Organizations such as the Science Commons are leading efforts to increase the availability and portability of scientific data [15]. The Science Commons is an offshoot of the Creative Commons project, which allows creators of intellectual works to establish terms for the re-distribution of their work that are much more generous than standard "fair use" protections. The Science Commons seeks to instill a similar spirit about the sharing of scientific data, with the goal of "accelerating the scientific research cycle."

Essentially, the data of interest to the Science Commons is grey content. Grey content can now be integrated into standard white literature, which is one reason why the distinction between grey and white literature is becoming moot.

The Canadian Institutes of Health Research (CIHR) released a policy in October 2006 that could greatly increase the prominence of health-related grey content in Canada. The "Draft Policy on Access to CIHR-funded Research Outputs" establishes several conditions that recipients of CIHR grants must meet. In addition to ensuring open access to peer-reviewed articles no later than six months after publication, CIHR grantees must provide public access to *research materials* and *research data* [16]. Many items qualify as research materials, including questionnaires, interview guides, and data abstraction forms. These are all examples of grey content. Research data encompasses "original data sets, data sets that are too large to be included in the peer-reviewed publication, and any other data sets supporting the research publication." The policy specifically encourages grantees to make their research data available in an electronic form. Although they are civil servants rather than open data activists, CIHR policymakers share the same motivation for increasing the availability of grey content.
The Canadian policy is a positive development, but activism about the value of open data remains necessary on an international level. Both the United States and European Union have passed legislation that makes it more difficult to share data, not easier [17, 18]. This is designed to protect the economic interests of data aggregators. Proprietary uses of scientific data are not always inappropriate, but should be on a "value-added" basis rather than through locking away raw data that is only available via a license or other means of payment.

Some scientists have taken their own steps to increase access to data. The journal *Nucleic Acids Research* (NAR) is now fully open access, after a year as a hybrid open access journal in 2004 and many years before that as a traditional subscription-based journal [19]. Electronic versions of NAR articles often contain "supplementary materials," which range from simple graphs to the more sophisticated grey

content of interest to open data advocates. For this paper, we evaluated whether NAR's move to complete open access produced a concomitant increase in the quantity and quality of the grey content integrated within NAR articles.

II: Case Study of *Nucleic Acids Research*

Nucleic Acids Research publishes articles about the "physical, chemical, biochemical, and biological aspects of nucleic acids and proteins" [20]. Its impact factor is in the top 10% of journals for Biochemistry & Molecular Biology, and has continued to rise since becoming fully open access. NAR articles are digitally archived in PubMed Central, beginning with the first issue published in 1974.

2002 and 2003 were the last years that NAR was published under a traditional subscription model. 2004 was a transitional year to hybrid open access, before full open access began in January 2005. NAR publishes 24 issues per year. Beginning with the first issue of 2002 until the 16th issue of 2006 (which was the most recent issue at the time of our study), we determined the percentage of "supplementary materials" that appeared in each issue. "Supplementary materials are denoted by a red flag appended to an article in PubMed Central. Not every article contains supplementary material, so the percentage equals the number of articles in each issue with supplementary material divided by the total number of articles in that issue. Yearly percentages are the average of each issue's percentage.

There has been a steady increase in the percentage of supplementary material published in NAR since 2002, with the exception of a modest decline from 2005 to 2006.

Year	Percentage of Supplementary Material
2002	11.27%
2003	16.36%
2004	22.16%
2005	31.17%
2006 (As of September)	26.42%

While the quantity of supplementary materials has increased, the caliber of these materials is of greater importance. Are they simply graphs bundled together as "supplementary," which are not any different from what you would find in print? Or are they qualitatively different examples of "grey content" that add value to the existing article?

For eight issues published in every year since 2002 (sixteen in 2002 and eight every year since then), we sampled five articles that contained supplementary materials. Using a scale from 0 to 2, we averaged the relative greyness of the supplementary materials. 0 = no difference from what you find in a standard journal article; 1 = some difference from the content in a standard journal article, and thus some greyness; 2 = a significant difference from a standard journal article, or the highest level of grey.

Using this scale, we found a modest increase in the caliber of grey content between 2002 and 2006. This is much less pronounced than the general increase in supplementary materials.

Year	Average Greyness (From 0 to 2)
2002	0.84
2003	0.99
2004	1.075
2005	0.75
2006	1.15

III: Conclusion—Toward Grey Content

The NAR case study reached a more modest conclusion than we had anticipated. While it would have been gratifying to proclaim a virtuous circle between full open access and enhanced access to grey content, the reality is that the quantity of supplementary materials increased much more substantially than the quality.

At the same time, it would have been much more sobering to report that the quality of grey content had *declined* since NAR became fully open access. NAR articles with quality grey content are examples of "datuments," a term coined by Peter Murray-Rust and Henry S. Rzepa in 2004 [21]. A datument is a "hyperdocument" capable of "transmitting and preserving the complete content of a piece of scientific work." By that definition, the NAR articles with top-quality grey content are certainly datuments.

The Internet is undergoing a profound transformation, from "a Web of connected documents to a Web of connected data" [22]. As this transformation unfolds, scholars of grey literature should shift their focus from managing discrete grey documents to curating diffuse grey content. The challenge of harnessing

such content is enormous, but worth the effort. Grey content is the foundation of scholarship, and we have an opportunity to make it much more accessible than ever before.

References

1. SHERPA/RoMEO-Publisher copyright policies & self-archiving. http://www.sherpa.ac.uk/romeo.php [Accessed 11 November 2006].
2. arXiv.og e-Print archive. http://arxiv.org/ [Accessed 11 November 2006].
3. Bethesda Statement on Open Access Publishing. http://www.earlham.edu/%7Epeters/fos/bethesda.htm [Accessed 11 November 2006].
4. Eysenbach G. (2006), Citation advantage of open access articles. In: PLoS Biology, 4 (5): e157, May 2006. Online at: http://biology.plosjournals.org/perlserv/?request=index-html&issn=1545-7885 [Accessed 11 November 2006].
5. Willinsky J. (2003), Scholarly associations and the viability of open access publishing. In: Journal of Digital Information, 4 (2): Article 177, April 2003. Online at: http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Willinsky/ [Accessed 11 November 2006].
6. Prosser DC. (2004), Between a rock and a hard place: the big squeeze for small publishers. In: Learned Publishing, 17 (1): 17-22, 2004. Online at: http://eprints.rclis.org/archive/00000945/ [Accessed 11 November 2006].
7. Hunter K. (2004), Open access: yes, no, maybe. Online at: http://www.nature.com/nature/focus/accessdebate/3.html [Accessed 11 November 2006].
8. Suber P. (2006), Nine questions for hybrid journal programs. Online at: http://www.earlham.edu/~peters/fos/newsletter/09-02-06.htm#hybrid [Accessed 11 November 2006].
9. Ibid.
10. Senate of the United States (2006), Federal Research Public Access Act of 2006. Online at: http://cornyn.senate.gov/doc_archive/05-02-2006_COE06461_xml.pdf [Accessed 11 November 2006].
11. ATA-Federal Research Public Access Act of 2006. http://taxpayeraccess.org/frpaa/index.html [Accessed 11 November 2006].
12. European Commission (2006), Study on the economic and technical evolution of the scientific publication markets in Europe. Online at: http://ec.europa.eu/research/science-society/page_en.cfm?id=3185 [Accessed 11 November 2006].
13. Harnad S (2006), First Things First: OA Self-Archiving, Then Maybe OA Publishing. Online at: http://openaccess.eprints.org/index.php?/archives/155-First-Things-First-OA-Self-Archiving,-Then-Maybe-OA-Publishing.html [Accessed 11 November 2006].
14. Banks MA (2005), Towards a continuum of scholarship: the eventual collapse of the distinction between grey and non-grey literature. Proceedings GL7: Seventh International Conference on Grey Literature, 2005. Online at: http://eprints.rclis.org/archive/00005803/ [Accessed 11 November 2006].
15. Science Commons. http://sciencecommons.org/ [Accessed 11 November 2006.
16. Canadian Institutes of Health Research (2006), Draft Policy on Access to CIHR-funded Research Outputs. Online at: http://www.cihr-irsc.gc.ca/e/32326.html [Accessed 11 November 2006].
17. United States Copyright Office Summary (1998), Digital Millennium Copyright Act of 1998. Online at: http://www.copyright.gov/legislation/dmca.pdf [Accessed 11 November 2006].
18. European Parliament (1996), Directive 96/9/EC on the Legal Protection of Databases. Online at: http://europa.eu.int/ISPO/infosoc/legreg/docs/969ec.html [Accessed 11 November 2006].
19. Suber P. (2006), Nine questions for hybrid journal programs. Online at: http://www.earlham.edu/~peters/fos/newsletter/09-02-06.htm#hybrid [Accessed 11 November 2006].
20. Oxford Journals-Life Sciences-Nucleic Acids Research. Online at: http://www.oxfordjournals.org/nar/about.html [Accessed 11 November 2006].
21. Murray-Rust P, Rzepa HS. (2004), The Next Big Thing: From Hypermedia to Datuments. In: Journal of Digital Information, 5 (1): Article 248, March 2004. Online at: http://jodi.tamu.edu/Articles/v05/i01/Murray-Rust/ [Accessed 11 November 2006].
22. Markoff J. (2006), Entrepreneurs See a Web Guided by Common Sense. In: New York Times, A1, November 12, 2006. Online at: http://www.nytimes.com/2006/11/12/business/12web.html?hp&ex=1163394000&en=a34a6306f48166fb&ei=5094&partner=homepage [Accessed 12 November 2006].

From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource Migration in European Context

Joachim Schöpfel, Christiane Stock, and Nathalie Henrot
INIST-CNRS, Institute for Scientific and Technical Information, France

Abstract
In 1980, some major European scientific information centres established the "System for Information on Grey Literature in Europe" (SIGLE) to provide access to European grey literature and to improve bibliographic coverage. August 23, 2006, the Luxemburg Register of Commerce and Societies published the liquidation of the association EAGLE that produced the SIGLE database until 2005.

Nevertheless, the former EAGLE member consented to preserve the European co-operation for grey literature and to transform the 1980 model into a sustainable network in the emerging environment of open access to scientific information.

The first step was to archive the SIGLE records in an open and freely searchable database, conform to the OAI metadata harvesting protocol. The French INIST developed OpenSIGLE based on MIT software (DSpace) and loaded most of the SIGLE records in a simplified XML format.

The communication provides an overview of the short history of SIGLE and EAGLE and describes how this unique resource was moved from a traditional host to an open access environment, giving the database a new look while preserving essential features characteristic for SIGLE.

1. EAGLE, a short history

In 1980, some major European scientific information centres established the "System for Information on Grey Literature in Europe" (SIGLE) to provide access to European grey literature and to improve bibliographic coverage (see Wood & Smith 1993).

The SIGLE database covered all scientific domains, STM, social sciences and humanities. In 2005, it contained 855,260 records from 16 countries and the European Commission; UK, German, French and Dutch records represented 90% of the overall input. 63% of the records were reports, 32% were theses and dissertations, and the other records were conference proceedings, data files and translations.

From 1980 to 1985, SIGLE was funded by the Commission of the European Communities (CEC). When CEC financial support ended in 1985, the national centres formed a network for the acquisition, identification and dissemination of grey literature called "European Association for Grey Literature Exploitation" or EAGLE, who became the producer of the SIGLE database.

EAGLE was created as a non-profit making association situated in Luxemburg. In the beginning, membership was limited to the member countries of the European Union (former European Community) but this limitation was abolished in 1995. In the end, 14 countries participated actively, and discussions on partnership were going on with other potential members, especially with East European and North African countries.

The economic model of the association was based on initial funding by the European Commission, from 1985 onwards on membership fees and royalties from servers and products (CD-ROM, records). The SIGLE database was distributed by BLAISE, STN, EI and Ovid (Silverplatter) and in the early nineties by SUNIST in France. Records were also sold to organizations like NERAC for specific use.

Costs were generated by the management of the association and mainly by the operating agent, whose task was to merge the national files into a unique server file every month.

2. Challenges and dead-ends

Twenty years after its creation, EAGLE faced four major challenges:

Internet and new technologies of information and communication: SIGLE offered no solution for online cataloguing, metadata harvesting, links to full-text and other resources.

Organisational structure: EAGLE was unable to take important technical and organisational decisions. Main members resigned from the association or intended to do so.

Coverage: National input became increasingly unrepresentative of the national production of grey literature, and input was continuously decreasing. Electronic documents were not referenced in the database.

Economic model: Investment for the development of the database was not provided.

EAGLE members were aware of these challenges and discussed possible options and solutions. Nevertheless, the organisational structure and the economic model of the non-profit making and low budget association did not allow necessary strategy decisions. It also became obvious then that no institutional member would invest more than its membership fee and that another funding from the EU would be unlikely.

3. Liquidation of EAGLE

In the face of this situation, the 2005 General Assembly from March 14, 2005, held at Karlsruhe, Germany, resolved unanimously to liquidate the association. The President (J. Schöpfel, INIST) and the Vice-President (S. Rehme, FIZ Karlsruhe) were appointed as liquidators.

After satisfaction of claims the remaining funds were distributed on equal parts amongst the members, taking into account the current memberships and the payment of the membership fees during the last years.

The contracts with STN, EI, NERAC and Ovid (SilverPlatter) were cancelled. The EAGLE website was deleted.

All usage rights of the SIGLE database lapsed upon complete liquidation of the association while the copyright on input remained with each member who supplied the records. Nevertheless, the General Assembly asked the operating agent, the FIZ Karlsruhe, for interim conservation of the SIGLE records in XML format beyond the liquidation of SIGLE, for the purpose of archiving and integration into a new European non- profit project. Nearly all of the former EAGLE members signed a declaration of intention regarding this future use of their existing input of the SIGLE database.

The complete liquidation of EAGLE was published by the Luxemburg Register of Commerce and Societies at August 23, 2006.

4. Perspectives and projects beyond EAGLE

At the same General Assembly in 2005, we presented the “MetaGrey Europe” project (see Schöpfel 2006). The objective was to preserve the European co-operation for grey literature and to transform the 1980 model into a sustainable network in the emerging environment of open access to scientific information, especially in the context of the 2003 Berlin Declaration.

The conditions were: no or low budgeting, investment if possible only through direct and non-financial contribution by participating structures (human resources…), no or small-scale organisational structure. EU funding could be helpful (we investigated EU Framework Programmes FP6 and FP7 options) but would generate too much executive work.

The first step was to archive the SIGLE records in an open and freely searchable database, compliant with the OAI metadata harvesting protocol. This part of the project is nearly finished. The French EAGLE member INIST developed OpenSIGLE based on MIT software (DSpace) and loaded most of the SIGLE records in a simplified XML format.

The second step is the identification and developing of European OAI initiatives for grey literature. Most of the former EAGLE members already are involved in such projects, especially for electronic theses and dissertations and reports, or intend to do so (see for instance Boukacem-Zeghmouri & Schöpfel 2006 or Stock, Rocklin & Cordier 2006). What is needed is an inventory of these projects and if necessary, technical and organisational assistance.

The third and most important step will be the development of a gateway to European grey literature. Last year in Nancy, we presented this project as a meta-search engine (“MetaGrey Europe”). The underlying idea was the indexing of European grey collections (catalogues, databases, full-text servers and archives) by a performing search engine that copes with different formats and document types, interdisciplinary and multilinguistic terminology. Whether we can co-operate with an existing search engine (Google, Scirus, Exalead) or need to develop a new search tool, perhaps together with the French-German Quaero project or the German Search Engine Lab at the university of Hannover, needs investigation and discussion.

These two steps will be our challenge for the coming years.

5. More information about the SIGLE database

Before talking about the migration from the host-based database SIGLE to OpenSIGLE, allow us to recall some facts and characteristics concerning SIGLE.

For each SIGLE member country, one or two national structures assumed the acquisition, referencing and document delivery of grey literature, mostly national libraries (UK, Luxembourg) or documentation centres of national research organizations (Italy, France, Germany). Merging of the national files was done by an independent operating agent under contract and the database was hosted on up to three different servers. A CD-ROM was produced from 1992 on by Silverplatter/Ovid.

Some characteristics

Cataloguing rules and classification scheme were adopted from the INIS database produced by the IAEA (International Atomic Energy Agency). The SIGLE classification is derived from the American COSATI scheme. The same scheme is today used by GreySource (GreySource 2006). One important difference to INIS is that SIGLE never included serials (only collections of monographs) and never provided records on an analytical level.

Each national structure sent records in its own language. A search through the entire database was made possible through providing an English translation of the title or English keywords. One of these fields was mandatory. This constraint may seem irrelevant from an American point of view, but for some countries it was a considerable barrier to increase input.

One of the goals of SIGLE from the beginning was to facilitate access to grey documents in Europe:

- Each record contained a clear mention of availability (with or without shelf number).
- Each member country committed to supply the referenced document on demand, whether from its own collections or through interlending service by back-up libraries.
- Useful information on document supply (addresses, conditions) was given on specific help pages or in user-guides.

Contents

SIGLE was started in 1980, but since some members put in older documents, the earliest publications go back to the sixties.

Although SIGLE was a multidisciplinary database, about one third of the records were found in the humanities and social sciences (>35%), followed by medical and life sciences (12%) and physics (9 %). These figures were related to different practices of research communities to publish white only or grey, but also to the willingness of organizations to cooperate. Some of the documents existed only in three copies and it was not always easy for the SIGLE centre to obtain one of them.

Among the document types one finds a majority of reports, followed by doctoral theses or dissertations. In several countries monthly files were obtained by conversion from other catalogues, without a clear identification of the document type, so that the "Miscellaneous" category is a "hold all" for these cases.

Over the years the content of the database improved. For instance, since 1997 English abstracts were added, in particular to Russian records. Members provided English keywords with increasing frequency. Even so, another important project was never realized: the integration of electronic documents, if possible with a link to the full text. Instead members started to build institutional repositories or to provide access to electronic grey literature by other means.

6. OpenSIGLE - moving the resources

Preliminaries

Once INIST decided to continue to make SIGLE available and to allow free access, we looked for the tools.

Throughout frequent visits to institutional or other repositories we had observed that not only non-textual documents became more frequent in these repositories or archives, but also that an increasing number of them provided only bibliographic records, without a link to the full text. This is particularly true for sites where the records come from catalogues, and the full text is added eventually (see Groningen 2006). DOAR and ROAR (see DOAR 2006 and ROAR 2006) provide useful information as to what kind of

information awaits you in a given archive or repository. So why not use an open source platform for SIGLE?

INIST had previously adopted the DSpace software for 2 in-house applications: I-Revues, a platform for open access journals and conferences (http://irevues.inist.fr), and LARA, a repository for grey reports, which was presented at GL7 (http://lara.inist.fr).

So we were already familiar with this platform, its advantages and its limits and constraints. Particularly useful in this case, we had acquired a first experience through the two aforementioned applications concerning a mass conversion and mass upload of data into DSpace. FIZ Karlsruhe, the last SIGLE Data Processing Centre (DPC) in office, had provided us with records in different formats, including XML, for evaluation.

A quick comparison of the metadata formats and indexes showed that it should be possible to move the SIGLE data to this software without too great a loss in information details.

The next step was the writing of specifications for the metadata conversion and the customization of DSpace. We also studied closely the old CD-ROM version of SIGLE in order to include a maximum of existing functionalities and indexes as well as online-helps in OpenSIGLE.

We were fortunate to have a student in computer science, specialized in Web applications, A. Adlani, as an intern for 10 weeks. With his help and invaluable suggestions, we were able to realize a prototype close to the final version.

Metadata conversion and migration

DSpace uses a qualified Dublin Core metadata set which is less detailed than the SIGLE metadata we received from the DPC (see DSpace Metadata 2006). FIZ Karlsruhe provided us with records in XML written in the SIGLE format and completed by some server-related fields. Therefore mapping was relatively easy, except for some coded fields and specific information. For example we wanted the English title to appear systematically in the field labelled "Title". In the source record it could be either in the field for the original title or in the field for the English title. Several specific fields from the source format were merged to one field for OpenSIGLE. Others were defined differently to fit with the metadata set. Some qualified fields were added to the metadata set used by DSpace without disturbing the OAI compliance: conference title, report number and availability statement.

The most significant change was a simplification in the document type information. The original SIGLE format distinguished between document type and literature indicator, but diverging conversion practices led to inconsistencies. In OpenSIGLE we propose a simplified list of the principal document types.

For the tests we used the French data and a selection of 1400 records from different countries provided by FIZ Karlsruhe. Several test uploads showed us inconsistencies in the display of data, which we tried to amend as far as possible. Those differences were due to either an "evolution" of some elements over the years or a change of the rules such as the constraint to provide an English title. This is why some recent records showed no information at all in the title column in the results display. Instead it now mentions "No title". Some fields in the source records came in a coded form (language, subject categories and document types). We tried as far as possible to convert them into full wording (e.g. inconsistencies for the language codes). We did not try to amend information which came with typing errors in the input files such as publication dates like 3001.

7. OpenSIGLE on DSpace: new presentation and new look

DSpace allows organizing the contents of a repository according to communities and collections. INIST decided to use 2 types of communities: the member countries and SIGLE subject categories on their primary level. Each country or subject category holds a collection of records. Some minor and less used subject categories were regrouped in one collection. In a mass upload on DSpace each records (or item) can be "attributed" to only one community or collection. We decided to choose the first classification code of each record. Since the files of each member country are treated separately, it is possible to declare also the country community for each record.

Contrary to the CD-ROM version the document type is no longer searchable in OpenSIGLE. We found it interesting to display the information in the list of results, along with the title, the authors and the publication date. This is not a feature of the basic version of DSpace, but we observed similar practices in other repositories (see ERA 2006 and Glasgow 2006).

The SIGLE classification scheme with its 246 subject sub-categories can be searched through the subject field, either by its code or its wording. A specific help page accessible at any moment lists the complete

classification schemes with both the codes and their description. As mentioned above, the subject areas were reduced to 15 entries for the organization of the database in collections and for browsing purposes.

For OpenSIGLE we chose the latest stable version available of the software which was then DSpace 1.3.2. One of the new features in this version is the support of multilingualism of the user interface (cf. DSpace system documentation 2006) This feature has been developed a little bit further by our student and OpenSIGLE can be used with interfaces in English (which is the main version), French, German and Italian, the four most representative languages for the database. The help pages and the "About" information are available in English and French only, since they must be translated specifically.

Document supply being very important for the SIGLE database, we decided to add an order form to facilitate the contact of the holder of the document or ex-EAGLE member. Of course the information about the document availability in each record was maintained, in a specific field. In addition we give updated information for each participating centre on the pages presenting the countries (in the "Countries" collections).

Last, but not least we customized the user interface with the help of our student: look, colours, size etc.

8. OpenSIGLE: results and outlook

The new OpenSIGLE repository gives open access to the bibliographic records of the SIGLE database from the members who signed the agreement.

With the migration to the DSpace platform look and presentation have changed.

Some data like the language or the document type are no longer searchable, but are still displayed, even in the list of results. The principal characteristics of the SIGLE database have been preserved or even improved. Access to the full text will be facilitated through an order form for document delivery and for some records hopefully through links to the electronic version in the future. Since the records are organized in collections based on the subject categories, and the OAI protocol for metadata harvesting considers collections as sets, a selective harvesting by subject will be possible.

More generally, OpenSIGLE seems to be the first migration of an important traditional bibliographic database into an OAI compliant environment. Some factors facilitated this migration: For the mapping of the metadata we had the advantage to pass from a very detailed format to a simpler one. The whole project benefited largely from our previous experience with DSpace and in particular from our knowledge about the import of records. Still OpenSIGLE provided us with a new experience concerning mass uploads on an Open Source platform; it is probably the most important upload for a DSpace repository ever done.

Although no longer updated OpenSIGLE remains a useful source identify and access to European grey literature. Three future developments may improve the visibility and usability of OpenSIGLE:

Links to the full text: Even if the new repository contains only bibliographic records, links from the OpenSIGLE metadata to the electronic full text where available are technically possible but have to be provided by the former EAGLE members.

Completion of content: Hopefully, those of the former EAGLE members who didn't sign the declaration of intention yet may reconsider their position and agree to the import of "their" national SIGLE input into the new database.

Referencing: Linking the OpenSIGLE records to scientific or general search engines, as a part of the MetaGrey Europe project, would largely enhance the visibility of the European grey documents of the last 20 years.

Acknowledgments to Anouar Adlani, a master student in information sciences from the Nancy university, and to Patrick Kremer from the INIST information system department for their helpful ideas and assistance during the preparation and implementation of the OpenSIGLE database.

Bibliography and resources

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Berlin 2003. http://oa.mpg.de/openaccess-berlin/berlindeclaration.html

Boukacem-Zeghmouri, C. & Schöpfel, J.: "Document supply and open access: an international survey on grey literature". – In: *Interlending & Document Supply* 2006, 34, 3, 96-104.

DSpace, a digital repository system, [homepage] (2006): http://www.dspace.org/ (accessed November 2006)

DSpace Metadata (2006): Metadata schema used by DSpace http://dspace.org/technology/metadata.html DSpace currently uses a qualified version of the Dublin Core schema based on the Dublin Core Libraries Working Group Application Profile (LAP). Some qualifiers were also added to suit DSpace needs.

DSpace (2006): Dspace System Documentation. http://www.dspace.org/technology/system-docs/history.html Version History/ Indicates as changes for DSpace 1.3 : Initial i18n Support for JSPs - Note: the implementation of this feature required changes to almost all JSP pages (accessed in November 2006).

ERA [homepage] (2006) : Edinburgh Research Archive : http://www.era.lib.ed.ac.uk/index.jsp a digital repository of research output from The University of Edinburgh, especially : http://www.era.lib.ed.ac.uk/browse-title (accessed in November 2006).

Glasgow [homepage] (2006) : Glasgow DSpace Service : https://dspace.gla.ac.uk/index.jsp The Glasgow DSpace Service is a repository for digital content including working papers, technical reports, theses and pre-prints by members of the University of Glasgow. https://dspace.gla.ac.uk/browse-title (last visited November 15th 2006)

GreySource (2006): http://www.greynet.org/greysourceindex.html : GreySource ~ A Selection of Web-based Resources in Grey Literature (accessed in November 2006).

Groningen [homepage] (2006): Dissertations of the University of Groningen http://dissertations.ub.rug.nl (accessed in November 2006).

I-Revues [homepage] (2006): http://irevues.inist.fr (accessed in November 2006). Website providing access to full text of serials and conferences.

LARA – Libre Accès aux Rapports [homepage] (2006): http://www.lara.inist.fr A federative repository for grey reports (accessed in November 2006).

OpenDOAR [homepage] (2006): The Directory of Open Access Repositories - OpenDOAR http://www.opendoar.org/index.html OpenDOAR is an authoritative directory of academic open access repositories (accessed in November 2006)

ROAR [homepage] (2006): Registry of Open Access Repositories (ROAR) http://archives.eprints.org/index.php (accessed in November 2006)

Schöpfel, J.: "MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE". – In: Farace, D. & Frantzen, J. (ed.): *GL7 Conference Proceedings. Seventh International Conference on Grey Literature: Open Access to Grey Resources.* Nancy, 5-6 December 2005. - Amsterdam: TextRelease, 2006.

Stock, C., Rocklin, E. & Cordier, A.: "LARA – open access to scientific and technical reports". – In: Farace, D. & Frantzen, J. (ed.): *GL7 Conference Proceedings. Seventh International Conference on Grey Literature: Open Access to Grey Resources.* Nancy, 5-6 December 2005. - Amsterdam: TextRelease, 2006.

Wood, D.N. & Smith, A.W.: "SIGLE: A Model for International Co-operation". – In: *Interlending & Document Supply* 1993, 21, 1, 18-22.

Knowledge Generation in the Field of Grey Literature: A Review of Conference-based Research Results

D.J. Farace and J. Frantzen, *GreyNet, The Netherlands*
J. Schöpfel and C. Stock, *INIST-CNRS, France*

Abstract
Perhaps the most cost effective research carried out in information science today is in the field of grey literature. The published proceedings in the GL-Conference Series provide a record of impressive research results in this field of information. These results are based on various and diverse approaches and methodologies, whereby citation data, survey data, systems data, bibliographic and metadata, as well as other evidence based variables and indicators are compiled, processed, and analysed. Notwithstanding the fact that knowledge generation is an important measure of wealth in science and technology, the costs in human and material resources appropriated from research budgets must also weigh-in to determine real effectiveness. Only in this way can our initial rhetoric be transformed into a working hypothesis. This research project, which lies within the structure of the GL-Conference Series, seeks to analyse not only the benefits of research on grey literature but also the costs related to carrying-out and publishing research results. In order to gather evidence-based data, a Project Information Document (PID) Form similar to others that are in place and use elsewhere has been designed to accommodate GL research. The PID-Form will be distributed to those authors/researchers, who respond to the GL8 Call for Papers, as well as to authors/researchers from last year's conference in this series.

Results gathered from these research resumes are expected to provide answers to relevant questions such as the percentage of research on the topic of grey literature that is formally funded, the ratio of ad hoc research, the ratio of individual to team research, average research costs and expenses, the duration of research projects, etc. This evidence-based data will then enable us to grasp the cost effectiveness of research on grey literature and compare other types of data compiled within a conference structure. And, in so doing, our results will help to demonstrate the power of grey literature to other information professionals as well as policy and decision makers, funding bodies and new investors. Furthermore, our results may be seen as indicative for other S&T conferences based on a call-for-papers.

Introduction

Knowledge generation is a process. Most often it is tedious work over a long period of time. Research results are like various sized blocks, hewn as it were and assembled to form a base. This paper presents an overview of conference based research results, which our team considers relevant to a better understanding of the field of grey literature.
This research project can also be seen as a follow-up to previous research carried out by the team over the past four years in which other types of conference based data were compiled via diverse methods, namely: a Review of the literature in 2003[i]; a General Survey on Grey Literature in 2004[ii]; Citation Analyses in 2004[iii] and 2005[iv]; and, an Author Survey in 2005[v].

Research Method

Practically, the same pool of respondents i.e. the Authors in the GL Conference series participated in all of the above-mentioned research projects. And, the same team of researchers *i.e.* project workers from INIST and GreyNet constructed and carried out the projects.

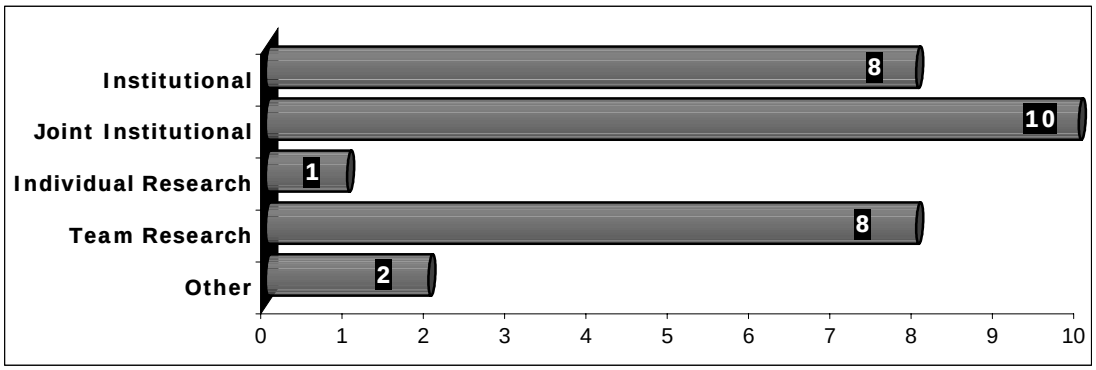
With the Call-for-Papers in 2005, those submitting proposals were asked to provide within the abstracts cost-related information on their research project. However, no standardized format was provided. Nineteen of the 29 submissions contained such information. The GL authors were thus prepared to provide this type of information in the same degree that they were willing to respond to the Author Survey *i.e.* roughly 56%. This then led us to undertake the present study.

We decided to implement a standardized form in 2006 that would better capture project data. A web search revealed that the U.S. Department of Agriculture[vi], the World Bank[vii] and IMF, as well as a number of CRIS, Current Research Information Systems, already have in place PID-Forms, which served as useful examples. We then drafted a PID-Form fitted to conference-based data. The form consists of 7 sections, totalling 20 response items, (***see Appendix***).

The PID forms were emailed to the first or sole author in each project both within the GL7 and GL8 Conference Programs. This totalled 52 projects of which 29 completed PIDs were returned and processed.

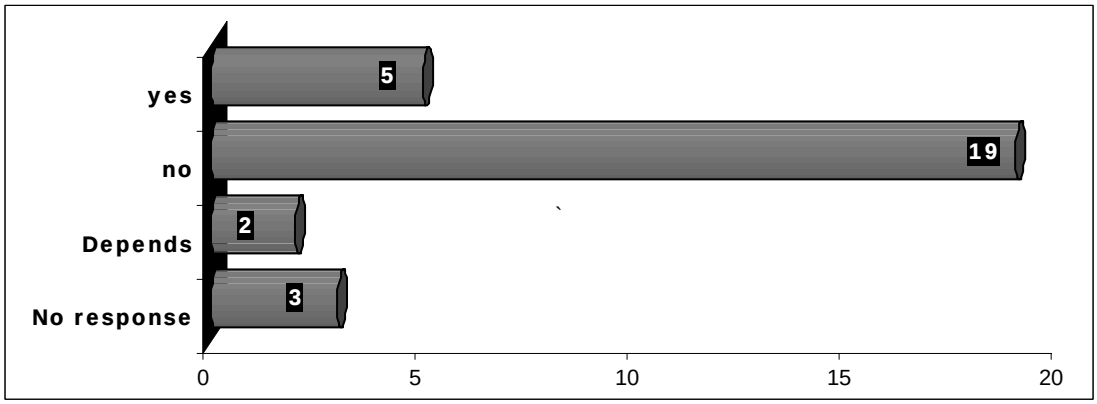
Project Information Document Results

Graphic 1 - Project Type



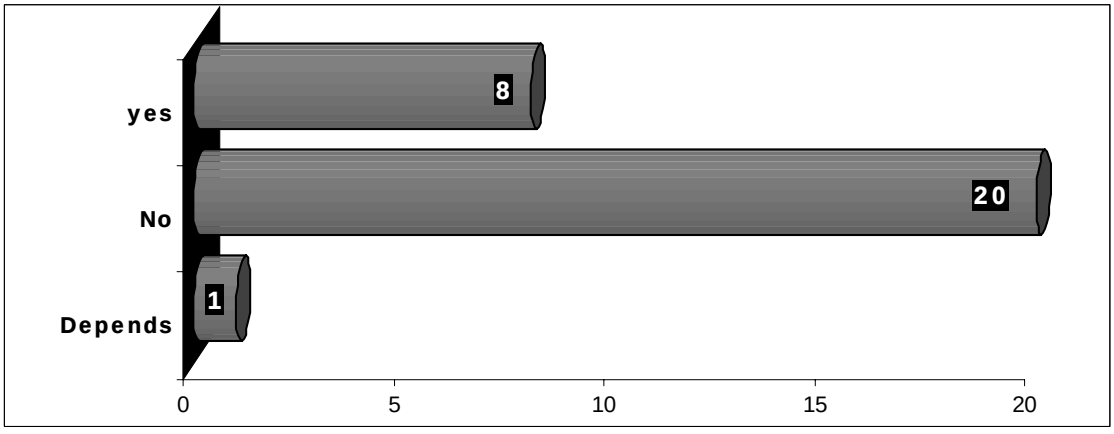
Graphic one illustrates the response to *item 1.2* on the PID Form, which deals with Project Type. Over a third (10/29) of the PIDs were JIRs, Joint Institutional Research Projects. When we look more closely at this type of project, we not only find that it was collaboration on an institutional level but also it was a cross-country or international collaboration. Fifteen years ago, research on grey literature was most often the work of lone individuals within a library or documentation center. Today that project type now appears to be the exception.

Graphic 2 - Required Publication



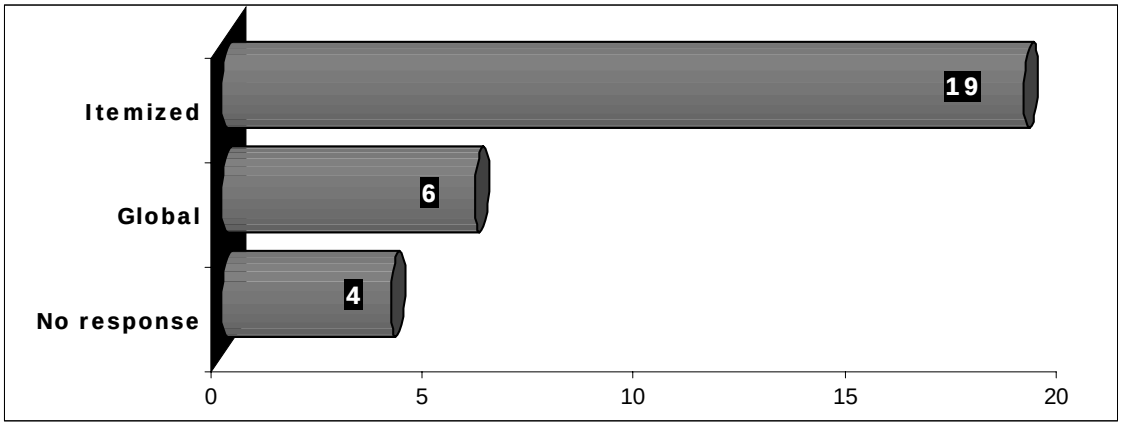
Item 1.3 on the PID Form inquires as to whether a project was linked to an academic title or requirement that a researcher publishes. Less than a fifth of the respondents (5/29) acknowledged that this was the case. And, while there is no previous data to draw comparisons, it would be worthwhile to examine future trends in this direction.

Graphic 3 - Grant/Funding Body



Graphic Three illustrates the response to *item 4.1* on the PID Form, which inquired whether funding or a grant organization was involved in the project? While less than a third (8/29) claimed this to be the case, half of those responses were Joint Institutional Research Projects (4/8). Which may lead us to believe that the best chance to receive project funding is within a JIR project structure.

Graphic 4 - Project Costs



Section 4 of the PID form allowed respondents to either itemize the project costs *(item 4.3)* or simply provide a global estimate of the project costs *(item 4.4)*. Our findings show that only a fifth of the respondents (6/29) did not provide a specific amount or sum in either of these response items.
Further, we can only assume that those who itemized the number of days invested in a project *(item 4.3.3)* did so by taking into consideration the number of project workers mentioned in *(item 2.1)*. Also we must assume that the average number of days in a working month is 22, whereby weekends are excluded. Future versions of the PID form should explicitly make reference to these points; and, in so doing reduce the number of assumptions made.

Table 1 - Uses and Applications

We see in this table the response to *(item 3.2)*, where the proposed uses and applications of research results were explained. Since this was an open question, the 29 possible responses had to be first coded and classified. This led to our six categories.

Uses and Applications	
Development of Resources	8
Theory and Knowledge	6
Quality Assessment	4
Management & Practice	8
Source of revenue	1
No Response	2

Table 2 - Project Duration

This table was calculated using the Project start date found in *(item 6.1)* and the Project termination date *(item 6.2)*. The average duration of a project was 23 months, which well exceeds the 9-month period between the Call-for-Papers and the presentation of the research results at the annual conference. The explanation of this should be found in the response to *(item 5.1),* Project History where prior or related research projects and proposals are documented. It is also worthwhile to mention that most of the research papers in the GL Conference Series relate to projects, which are not exclusively limited to grey literature.

Project Duration in months	
Average	23
Maximum	108
Minimum	1

Conclusions and Recommendations

The PID–form was distributed between 6 to18 months after the Call-for-Papers. Six months distanced from GL8 (December, 2006) and eighteen months distanced from GL7 (December, 2005). Had the researchers i.e. project workers received the PID-Forms immediately upon submission of their abstract, they may have been made more conscious of cost related issues in their projects and in turn the results would have been more comprehensive.

Interesting to note, a completed PID-form was regarded as a Project Information Document i.e. a specific type of grey literature, much like fact sheets, technical notes, tenders, etc. Also, in November 2006, during a series of guest lectures, a group of graduate students in information studies at the University of Amsterdam were asked to comment on our PID-Form used in gathering conference data. As newcomers to the GL Conference Series their interest was first of all with *(item 1.1)*, the Title of the Project and secondly with *(item 3.2)*, the Uses and Applications of the Results. I could understand their interest in the later, and came to realize that the project titles could also be coded and classified by subject area, giving an outsider group a clearer understanding of where research was being carried out in the field of grey literature.

As to whether our results may be seen as indicative for other S&T conferences based on a call-for-papers, the answer leans to the affirmative. Just recently, the National Library of Canada in Montreal requested a summary of conference registrations fees over the past five years. They want to organize an international conference in 2008. GreyNet has this information and is certainly willing to share it with other conference organizers, where knowledge generation in the field of grey literature stands to benefit.

In fine, one way of better sharing conference based results with other communities of researchers would be to make the PIDs, openly available on the conference site. A revised version of the PID-Form should clearly state that the information and data provided by the project workers would be publicly accessible. In fact, the revised PID-Form could replace the instrument used in previous Call-for-Papers simply by introducing in Section 3 of the PID Form, an item titled ‘Abstract’. And, during the course of their research, project workers could periodically revise and update their Project Information Document up until its termination date.

Appendix: PID – Project Information Document

GL Project Résumé Sheet

1.1. Title Project: *(If it carries the title of the Conference Paper enter "Same")*

1.2. Project Type: (*Indicate by placing an* [x] *in the appropriate box)*

- [ ] Institutional (only within your institution)
- [ ] Joint Institutional (in cooperation with other institutions)
- [ ] Individual Research (non-institutional)
- [ ] Team Research (collaborative, non-institutional)
- [ ] Other:

1.3. Is this project linked to an academic title, required publication, etc.?

2.1. Name(s) Project Worker(s):
2.2. Email Contact Address:

3.1. Approach/Methodology:
3.2. Proposed Use and Application of Results:

4.1. Funding/Grant Organization(s): *(If none skip to question 4.3)*

4.2. Project Budget: *(Total estimate in US$ or Euro)*

4.3. Project Costs and Expenses *(Itemized)*

- 4.3.1. Special equipment and supplies:
- 4.3.2. Salaries *(only if directly related to the project)*:
- 4.3.3. Estimate total N° of days invested in the project:
- 4.3.4. Travel, Lodging:
- 4.3.5. Conference Registration:
- 4.3.6. Other:

4.4. Project Costs and Expenses *(Total amount from 4.3 or a global estimate)*

5.1. Project History: *(Prior or Related Research Projects and Proposals)*

6.1. Project Start Date: (mm/dd/yy)
6.2. Project Termination Date: (mm/dd/yy)

7.1. Other Comments:

GL/PID-FORM/18AUG06

References

[i] Farace, D.J. and J. Frantzen (2004), Four Winds on the Grey Landscape: A Review of Four Information Professionals, their Work and Impact on the Field of Grey Literature. - In: GL5 Conference Proceedings. – Amsterdam : TextRelease, pp. 10-12. – (ISSN 1386-2316 ; No. 5)
ISBN 90-77484-01-9

[ii] Boekhorst, A.K., D.J. Farace and J. Frantzen (2005), Grey Literature Survey 2004: A Research Project Tracking Developments in the Field of GL. – In: GL6 Conference Proceedings. – Amsterdam : TextRelease, pp. 1-9. – (ISSN 1386-2316 ; No. 6)
ISBN 90-77484-04-3

[iii] Schöpfel, J., C. Stock, D.J. Farace, and J. Frantzen (2005), Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. – In: GL6 Conference Proceedings. – Amsterdam : TextRelease,
pp. 55-63. – (ISSN 1386-2316 ; No. 6)
ISBN 90-77484-04-3

[iv] Farace, D.J., J. Frantzen, J. Schöpfel, C. Stock, and A.K. Boekhorst (2006), Access to Grey Content: An Analysis of Grey Literature based on Citation and Survey Data: A Follow-up Study. – In: GL7 Conference Proceedings. – Amsterdam : TextRelease, pp. 194-203. – (ISSN 1386-2316 ; No. 7)
ISBN 90-77484-06-X

[v] Idem

[vi] http://www.csrees.usda.gov/business/businessfaqs/ad416r01.pdf
USDA Research Work Unit/Project Description (Accessed January, 2007)

[vii] http://www-wds.worldbank.org/external/default/main?pagePK=64187835&piPK=64620093&theSitePK=523679&menuPK=64187283&siteName=WDS&pageSize=20&docTY=540659
World Bank Project Information Documents (Accessed January, 2007)

Louisiana Coastal Wetlands and Louisiana Coastal Grey Literature: Vanishing Treasures

Gina R. Costello
Special Collections Library, Louisiana State University, 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 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

[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.

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

[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

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

Searching down the fisheries information web: An initial reflection

Janet Webster
Hatfield Marine Science Center, Oregon State University, United States

Patricia Merrikin and Jean Collins
Food and Agriculture Organization of the United Nations, 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.

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.

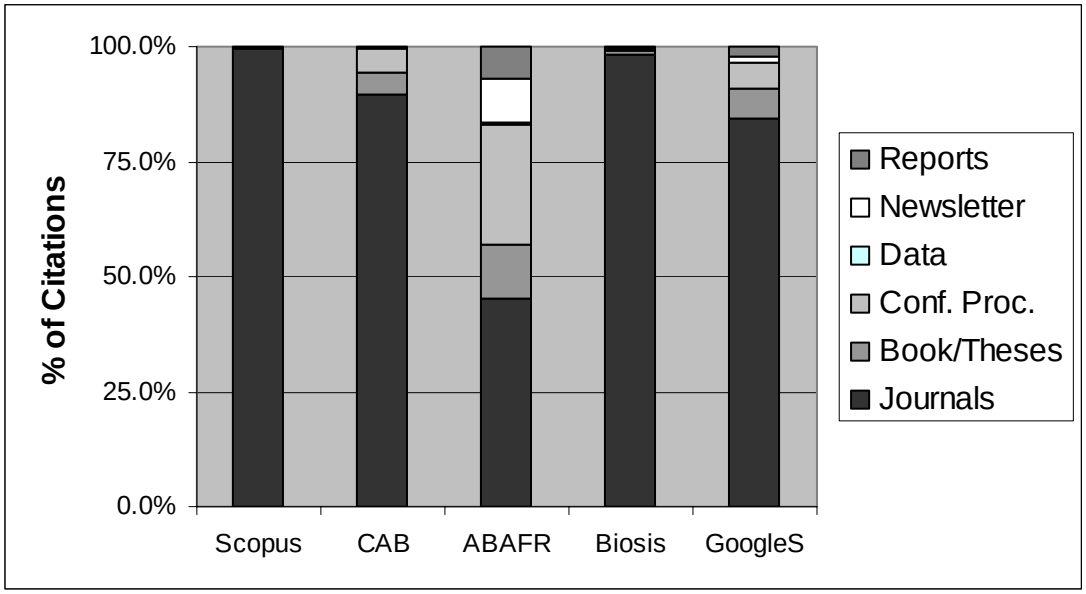
The following presents initial data from the analysis of all results by search tool. We also analyzed the data with duplicates removed, but the findings were similar. The results allow us to examine the efficacy and focus of various search tools. They also give a picture of who is publishing and where. 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	

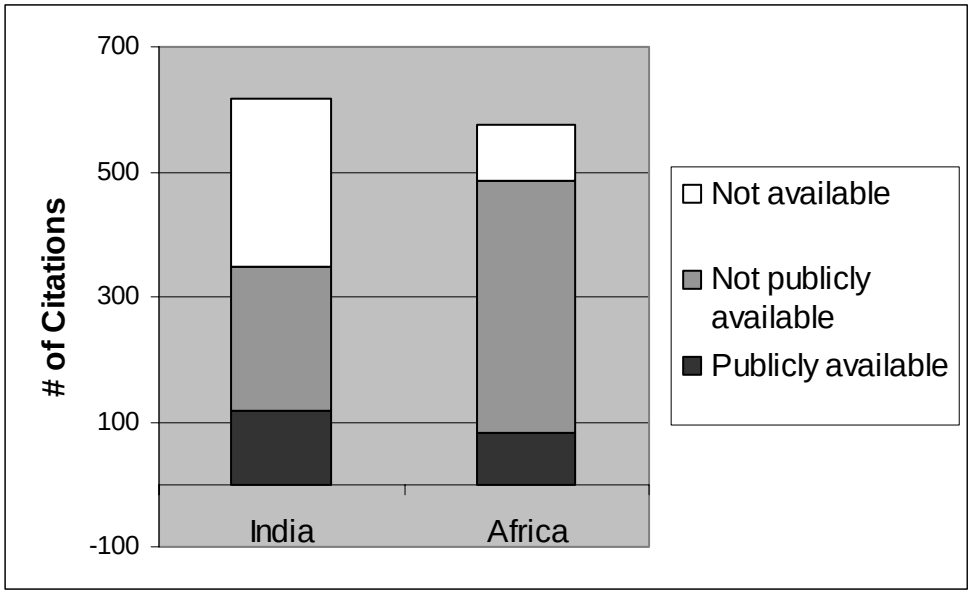
Table 9: Format - Africa

	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.

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, and Pete Reehling, *University of South Florida, United States*
Courtney Crummett, *National Library of Medicine, 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.

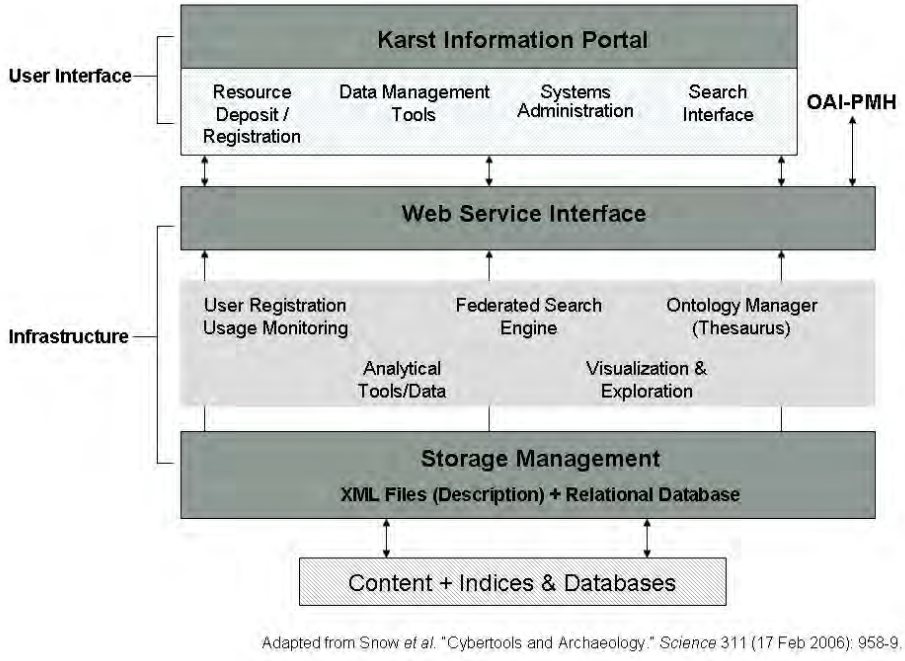


Figure 1. Graphic Representation of the Karst Information Portal.

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.

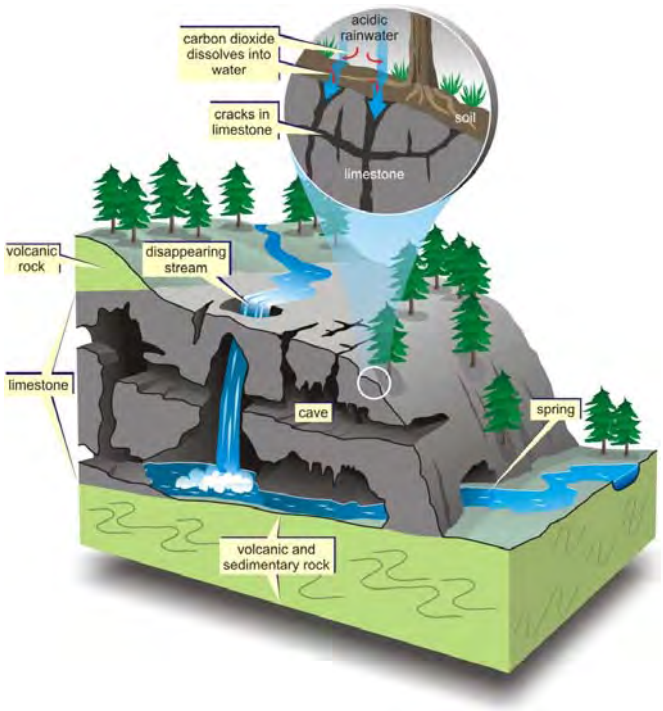


Figure 2. Karst Terrain (Natural Resources Canada).

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

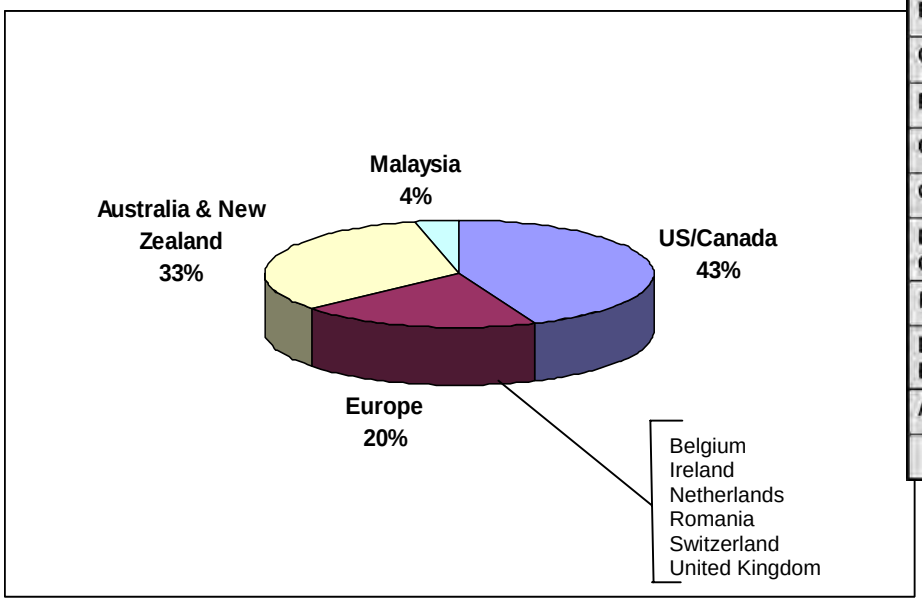
Information specialists from the Libraries and the School of Library and Information Science at the University of South Florida planned and conducted a global information needs assessment for the KIP. The survey was designed to elicit responses in three categories of information need: 1) information 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

Information Type (Listed in Order of Frequency of Use)	Responses (No.)		
	Use Type n=58	Produce Type n=52	Difficulty Locating Type n=49
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
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	

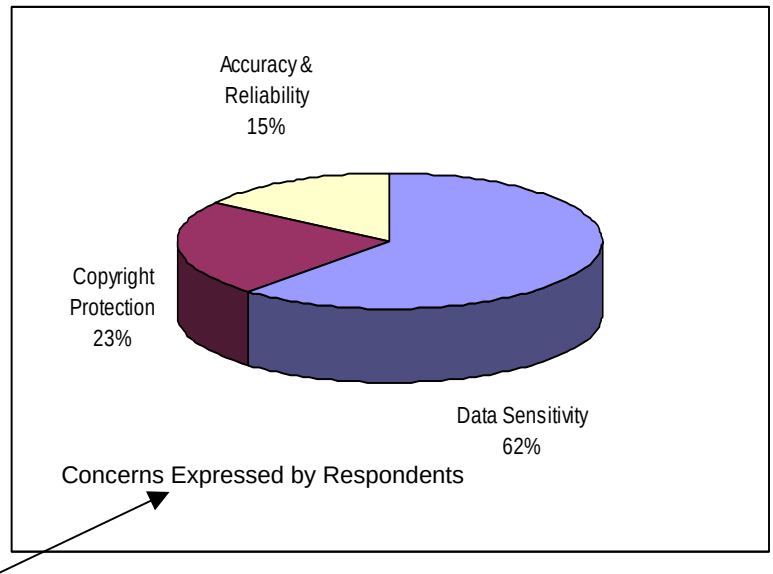


Figure 6. The KIP as Grey Information Repository.

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).

Assessing the Diffusion and Impact of Grey Literature Published by International Intergovernmental Scientific Groups: The Case of the Gulf of Maine Council on the Marine Environment

Bertrum H. MacDonald, Ruth E. Cordes, & Peter G. Wells
Dalhousie University, School of Information Management & School for Resource and Environmental Studies Canada

Abstract
Although many governmental and intergovernmental organizations publish vast quantities of grey literature, the importance of the diffusion and impact of this literature are rarely studied. Evidence from an investigation of the grey literature output of GESAMP, the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (sponsored by the UN and several of the UN-family of organizations), indicated that the literature reached scientific readers and was cited. To determine whether that evidence was representative of international intergovernmental bodies, another intergovernmental organization devoted to marine environmental issues, namely, the Gulf of Maine Council on the Marine Environment (GOMC) was studied. GOMC, an American-Canadian partnership, has been working since 1989 to maintain and enhance environmental quality in the Gulf of Maine. Through its own publications and others resulting from studies conducted under contract or in cooperation with other organizations, GOMC provides a complex publishing history for investigation. Over 300 publications were identified and over 500 citations were located after extensive searching using several citation tools. Citation patterns for GOMC publications mirror the findings of the study of GESAMP; grey literature is cited over lengthy periods, but grey literature tends to be cited primarily by other grey literature. Although digital alerting and access tools are increasing in number and coverage, a reliance on grey literature as the primary means of publication continues to pose hurdles for influencing scientific research, public policy, and public opinion. While grey literature is common to organizations such as GOMC and GESAMP, the impact of this literature can be muted because of the limitations of its dissemination and perceptions of its quality.

Acknowledgements: Support for this study was received from the Gulf of Maine Council on the Marine Environment, Environment Canada, and the School of Information Management, Dalhousie University. Patricia Hinch (Nova Scotia Department of the Environment and Labour), Michele Tremblay, David Keeley, and Peter Taylor (Gulf of Maine Council on the Marine Environment), and numerous librarians in the Gulf of Maine region offered valued advice and assistance.

Dedication: This paper is dedicated to the memory of Susan Snow-Cotter, who passed away December 2006. She was a long-term member of the Working Group of the GOMC, a member of Council, and a person devoted to wise environmental management of the Gulf of Maine.

Introduction

It is widely recognized today that environmental issues "have expanded from local and regional problems...to complex, interactive, and persistent problems that threaten the planet as we know it" (Doern & Conway, 1994, p. 4). Over the past quarter century governmental research units, university departments, and many other public and private sector organizations have devoted extensive resources to scientific and social science research on environmental subjects. These initiatives have prompted local, regional, national, and international governmental agencies to produce thousands of reports on environmental topics, most often published as grey literature. Major publications, such as the *Stern Review on the Economics of Climate Change* (Stern, 2006), released in the UK in October 2006, have received broad international attention (BBC News 2006a & b; CBC News, 2006). The same occurred for the vitally important *Millennium Ecosystem Assessment Reports*, the first of which was published in March 2005 (Millennium Ecosystem Assessment, 2005). While those particular reports may in fact foster public policy debate and action (it is too early to tell, if that will be the case), what is the fate of many, many other documents that fall below the radar of national and international media? Even when the reports are of the stature of the *Stern Review*, do they reach readers when and where it matters or are they only noted for short periods of time?

It is clear that the production of grey literature is central to the publishing practices of many organizations (e.g., O'Dell, Dallman, Vesely & Vigen, 2003; MacDonald, Cordes & Wells, 2004). But is that literature found easily when needed and used to maximum efficiency? Are the extensive resources devoted to the production of such publications justified (in some instances upwards of $1 million per title)? Are there better ways of ensuring the important scientific and technical assessments found in grey literature come to the attention of policy makers, stay in their view, and are used when appropriate to guide much needed environmental policies and other actions? Given the urgent nature of a number of environmental issues, such as climate change, these questions are not trivial, or easy to answer (Acreman, 2005; Francis, Whittaker, Shandas, Mills & Graybill, 2005; Schrecker, 2001; Roux, Rogers, Biggs, Ashton & Sergeant, 1999).

Such questions and others guided an investigation of the grey literature produced by GESAMP,[1] a significant international marine scientific advisory body (MacDonald, Cordes & Wells, 2004). That study demonstrated that marine environmental reports published as grey literature are used, often over lengthy periods, even though the publishing practices of GESAMP's sponsoring agencies have often hindered wide-spread distribution of its publications, and resulting awareness of its reports was often low. To test whether those findings are representative of major governmental and intergovernmental agencies focussed on environmental topics, an examination has been undertaken of the work of another intergovernmental body with multiple jurisdictional responsibilities, namely, the Gulf of Maine Council on the Marine Environment (GOMC) (www.gulfofmaine.org). The primary questions in this research are: 1) What has GOMC published, where and how? 2) What does citation analysis show regarding the distribution and use of GOMC publications? 3) What other approaches can be deployed to demonstrate the use and influence of GOMC publications? 4) What fundamental principles regarding the dissemination of grey literature from such organizations are emerging from this new study? and 5) Ultimately, is the Gulf of Maine better off environmentally from all this publication effort? Has the grey literature production had its intended effect on human behaviour towards the environment?

Mandate and History of the Gulf of Maine Council on the Marine Environment

The work of the Gulf of Maine Council on the Marine Environment began in 1988 with discussions on the need for a regional interagency organization, and initiation of a "State of the Gulf on Maine" report, which was released in December 1989 (Van Duesen & Johnson Hayden, 1989). The Council was established formally by the Premiers of the Canadian provinces of Nova Scotia and New Brunswick and the Governors of the American states of Maine, New Hampshire, and Massachusetts in December 1989 at an inaugural conference held in Portland, Maine. According to Allen L. Springer (2002), "the Council's creation responded both to the perceived need for increased institutionalization of patterns of cooperation in the North Atlantic region, and to the desire by state and provincial actors, both governmental and non-governmental, to play a more central role in that process."

The Council focuses on the marine environment of the Gulf of Maine and the Bay of Fundy. This marine region has a very high biological productivity and diversity, abundant and very valuable fisheries, habitats for numerous endangered or threatened species, many threats to its health and ecological integrity, and established economies and life styles linked to the sea in its many coastal communities. This focus also includes consideration of the land-sea interface, and the watersheds and estuaries of the Gulf region. The underlying philosophy in the Council's work is that activities on the land ultimately affect the sea, especially in more shallow coastal areas.

GOMC is an international intergovernmental body, with linkages to non-governmental organizations (NGOs), and the university research sector. Its focus is the marine environment of the Gulf of Maine, and environmental issues and their resolution, particularly ones of a cross boundary nature (e.g., air and water pollution, conservation of critical habitats and hemispheric migratory species, climate change, and introduced species). With the exception of a 2004 report, *Tides of Change Across the Gulf* prepared by Pesch & Wells, fisheries issues have not received extensive attention of the Council. The Council has had limited direct relations with citizen-based NGOs, and at times works with unpredictable finances received from its members. Some governmental agencies, particularly the U.S. National Oceanographic and Atmospheric Administration (NOAA) and the Environmental Protection Agency (EPA), have been the main fiscal supporters, although substantial financial and in-kind support comes from all members. GOMC's programs are financed with a budget normally in the $0.5-1.0 million per year range, excluding extensive in-kind support; matching dollars are in a ratio of 3 or 4 to 1. The Council works with refreshed five-year Action Plans, and is currently operating under its fourth Plan which spans fiscal years 2007-2012. The current Plan has three primary goals: habitat conservation and restoration, human and ecosystem health, and environmental sustainability.

Among GOMC's several long-term, flag-ship initiatives are its quarterly newspaper *The Gulf of Maine Times*, the Gulfwatch contaminants monitoring program, a salt-marsh restoration program, the Council's website (www.gulfofmaine.org) which includes services such as a "People Finder" and an "Inventory of Monitoring Programs," and an active publishing agenda. In addition, the Council maintains several funding programs for studies conducted externally (e.g., its external action grants), and a distinguished awards program to recognize outstanding achievements.

The Council Secretariat rotates among the five states and provinces on an annual basis, and is chaired by an individual in the host jurisdiction. The Council itself convenes twice a year, in one- or two-day meetings held in the host state or province, and is attended by political Cabinet Ministers, Deputy Ministers, Commissioners, or their representatives. The Council's mandate is carried out primarily through its Working Group, which reports to the Council, and is also chaired by a representative of the host jurisdiction. The Working Group holds quarterly two-day meetings, which are attended by senior policy managers and scientists directly involved in the Council's programs, and the Council's Secretariat. Several committees and subcommittees, which meet at least once per year, report to the Council's Working Group.

[1] GESAMP is formally titled IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection.

These committees, e.g., the Habitat, Monitoring, and Public Education and Participation committees, are co-chaired by American and Canadian members. The actual committee structures and agendas are steered by the action plans.

Overall, the GOMC's work entails research, ecosystem monitoring, communication and education, and public policy. Research is linked to and integrated through the Regional Association for Research on the Gulf of Maine (RARGOM), currently coordinated by the University of New Hampshire, Durham, New Hampshire, as well as through the facilities of the member institutions. Monitoring is conducted through habitat and contaminant subcommittees. The Council's significant communication agenda is pursued primarily through its website and publications, and many widely-attended workshops on a variety of topics (e.g., monitoring programs, salt marsh restorations, indicators of environmental change, and climate change). The Council also encourages public policy discussions, through both academic studies and public forums associated with Council meetings (e.g., wind farms, climate change, coastal zonation, and indicators for monitoring). The core work of GOMC is conducted with individual researchers and through the work plans of the member agencies. As shown below, the Council's work has been extensive in scope and prolific.

GOMC Publications

The Council has produced a large and diverse body of publications since it was established in 1989. The earliest items, an inventory of marine environmental quality monitoring programs in the Gulf of Maine and the first issues of the *Turning the Tide* newsletter, were published by the Gulf of Maine Working Group several months before the Council was officially created. The goal of many Council publications is to disseminate information to environmental managers and other decision-makers. As a result, visual appeal and clear language are prominent features of publications ranging from *The Gulf of Maine: Sustaining Our Common Heritage* (Van Duesen & Hayden Johnson, 1989), to the recent *Gulf of Maine Marine Habitat Primer* (Tyrrell, 2005).

In the seventeen years of its existence, the Council, working by itself or in collaboration with others, has published widely, including conference proceedings, technical reports, conference background documents, annual reports, action plans, newsletters, newspapers, magazines, fact sheets, brochures, maps in poster format, and a video. Other groups also have benefited from the Council's support for producing a similar array of publications. Moreover, individuals associated with the Council have given many workshop and conference presentations and written primary journal articles, resulting in another class of publications.

By the mid-1990s, the Internet had become a very important medium for communication and publishing. Many of the print publications that the Council had produced since 1997, along with a few from earlier years, are now available on the Council's website, most as easily printable PDF files (Table 1). The website is itself an evolving publication in its own right, presenting information in ways that are particularly adapted to the medium. For example, information from the GOMC's Gulfwatch monitoring program can be displayed with an interactive mapping tool, and the *Gulf of Maine Habitat Restoration Portal* offers information about the hows and whys of restoration projects. The new *KnowledgeBase* interface will soon allow searching for information in a variety of ways. The Council's website is taking on an increasingly important publication and communication role, but print publications remain an important part of the history and on-going work of the Council. Both print and digital publications currently have roles for information dissemination, but the Council's clear trend is to increasing emphasis on electronic media and its website.

Table 1: GOMC Publications* (Total 321)

Type of Publication	# items	# online
Council publications		
Action plans, work plans, annual reports (incl. 6 drafts)	27	6
Brochures and posters	13	4
Conference background papers	32	29
Conference reports	26	13
Fact sheets	10	7
Reports (incl. 6 interim reports)	85	31
Serials	4	1
Miscellaneous	17	6
Council-supported publications		
Conference reports and background documents	25	8
Other reports	13	7
Miscellaneous items	12	5
Reprints of Council authored documents (Mostly "Gulf of Maine Times" articles)	14	13
Publications based on Council work		
Journal articles	4	2
Papers in conference proceedings	9	2
Abstracts of conference presentations or posters	30	18

* Identified as of December 1, 2006

Early in its history the Council recognized that "in order to effectively manage the Gulf of Maine as the ecosystem that it truly is, decision-makers must have access to data and information from sources throughout the entire Gulf of Maine system" (Gulf of Maine Council on the Marine Environment, 1993, p. 4). An "explicit priority in the [first] Action Plan was the design of a computer system to manage large amounts of information about the Gulf of Maine ecosystem and to organize and package it in a way that would be usable by many audiences: state, provincial and federal agencies; laboratories; universities; schools; and non-governmental organizations....The long-term goal of the system is to allow people to exchange information, display graphics, search data and answer specific questions, and directly communicate with one another" (Gulf of Maine Council on the Marine Environment, 1996, p. 11). Publication and communication of information about the Gulf of Maine was a priority of the Council, and it has remained so.

Documenting GOMC publications

A number of valuable regional collections exist in New Hampshire, Maine, and Nova Scotia, but GOMC has not maintained a formal and comprehensive collection or list of its publications since it was established in 1989. Therefore, it was necessary to gather evidence of its publications from a variety of sources, including personal collections of GOMC publications and documents, the GOMC website, nearby libraries, other library catalogues, and web search engines. Evidence of GOMC-related journal articles and conference presentations were also found by searching article databases, electronic collections of proceedings, and print copies of other proceedings. These searches also led to the discovery of items published by other organizations with GOMC support. Some publications have effectively vanished, with electronic versions no longer available on the web and no print copies listed in library catalogues; only citations or mentions in other documents remain. While a large number of GOMC publications has been identified (Table 1), probably some have been missed, particularly if they were published during the Council's early years of operation. Some information about a collection at the Maine State Planning Office in Augusta is found on the GOMC website; if all items in that extensive collection were examined, additional early GOMC publications might be identified.

A database of records of the GOMC publications was created using ProCite software. Some draft versions of documents were included in the database, if copies were held by libraries or if they were cited. No attempt was made to record the detailed contents of the *Gulf of Maine Times* and the Council's earlier periodicals in the database. Many articles were published in the *Times* (currently 38 issues, 1997-), *Our Common Heritage* (2 issues, 1995-96), *Program Highlights* (30 identified issues, 1990-1996), and *Turning the Tide* (16 identified issues, 1989-1993). Documents such as press releases, briefing documents linked to the Council "meetings" webpage, or the individual presentations (PDFs of slides) from the Northeast Coastal Indicators Workshop, held in January 2004, were also excluded from the database (although they are available on the GOMC website).

Locating Citations to GOMC Publications

Since citations provide one indicator of the use of publications, searches of citation databases were undertaken to reveal which GOMC publications have been cited in the research literature. Environmental managers, who are the intended audience for many of GOMC's publications, are more likely to prepare technical reports than to write journal articles, so strategies were developed to locate citations in such reports published on the Web.

Web of Science, Scopus, Google Scholar and *Google* were used for citation searching in this study. In the study of GESAMP, completed in 2001-2002, Thomson ISI's *Web of Science* citation databases, which index a broad interdisciplinary collection of research journals dating back several decades, were the only tools for systematically locating citations. Those citation databases use a compressed citation format which works well for citations of journal articles but makes it difficult to locate citations of report literature. Since 2002, additional tools for citation searching have become available. Elsevier's *Scopus* database, launched in 2004, indexes a broader range of periodicals than *Web of Science*, but the currently indexed citations primarily cover articles published since 1996 (Elsevier, 2006). *Google Scholar* indexes journal articles and selected web-based resources, but a description of its coverage is not available. *Google Scholar* searches sometimes return relevant results from *Google Book Search*, so some citations in books were identified. Unlike *Web of Science,* both *Scopus* and *Google Scholar* provide the full text of each citation, making it possible to more accurately locate citations to technical reports.

Since web search engines such as *Google* and *Yahoo* now index PDF documents, *Google* was used to locate citations in reports published in that format, but such searching is not very efficient. In both *Google* and *Google Scholar*, each likely file must be opened and searched to determine whether it contains a reference to a specific GOMC publication, or merely a mention of the Council and its work. When a large report has been broken into several files and a citation is found in one of them, additional time is needed to identify the citing document by locating its first section or the linking webpage. Recording information about citing documents is much more time-consuming than importing information about citing journal articles from an article database directly into bibliographic management software. Notwithstanding difficulties in locating citations, the following discussion demonstrates characteristics of GOMC publications, and their use and influence.

Discussion of Findings

GOMC Publications

Since its creation in 1989, GOMC has produced over 300 publications, the majority of which were prepared by the Council itself (Table 1; Cordes, MacDonald & Wells, 2006). GOMC provides financial support to related organizations, which have published at least 50 reports and documents with GOMC sponsorship. Furthermore, researchers, managers, and consultants associated with GOMC through its Working Group or committees have published conference papers and papers, which draw on the work of the Council. While early GOMC publications were produced only in print, most items produced since 1997 have been available in both print and electronic format. For example, the most recent report, *Cross Border Indicators of Climate Change Over the Past Century: Northeastern United States and Canadian Maritime Region* (Wake, Burakowski, Lines, McKenzie, & Huntington, 2006) is available in both formats (with a limited print-run of the print version).

In contrast to the publishing pattern of GESAMP, which has produced about 115 publications, plus a similar number of translations or reprints in other report series, over almost four decades, GOMC has been more prolific. Whereas GESAMP publishes mostly major technical reports in its advisory capacity to its UN agency sponsors, GOMC's mandate includes responsibility for public education. Thus, GOMC places greater attention on design and readability. However, greater concern for accessibility has not translated into consistent attention to dissemination of publications, nor care in describing publications for identification and access. The multi-jurisdictional nature of GOMC and its rotating governance model has contributed to inequity in distribution of its publications. For example, when the *Gulf of Maine Marine Habitat Primer* was published (Tyrrell, 2005), copies were mailed to a large list in the United States, but distribution was much less effective in New Brunswick and Nova Scotia (Cordes, MacDonald, and Wells, 2006). Even when publication is in digital format, as in the recent report on the important topic of climate change (Wade, et al., 2006), significant details such as the date of publication are not obvious. Like many other organizations that rely on grey publications (whether in print or online), GOMC's interest is often focussed primarily on the content of documents rather than diffusion and accessibility matters. Once a work is published, attention moves rapidly to other projects rather than providing additional resources to ensure that the published work is disseminated (in the case of print copies) and effectively designed and described for searchability and heightened awareness (in the case of digital copies).

While GOMC has a specific mandate to integrate and communicate findings for regional environmental management of the Gulf of Maine, the Council has not until now maintained a comprehensive list of its own publications to document its output and influence (Cordes, MacDonald & Wells, 2006). This inattention to recording its past output is not uncommon among organizations of multi-jurisdictional structure. Publishing to the web has the potential of reducing this oversight, since publications can be maintained in continuous existence, and development of digital discovery tools might overcome the lack of comprehensive publication records.

Results of the Citation Analysis

Citation data confirm that GOMC publications are used worldwide, but primarily by authors within the region of the Gulf of Maine (Table 2). Over 500 citations — to publications in the first category of Table 1, and to four journal articles based on GOMC work, tracked through the citation search noted above — are related to GOMC's print and online publications. Within the Gulf region and in Canada outside of the Gulf region, authors show little difference in their preference for print or online publications, but beyond these two areas mostly digital copies of publications are cited. This citation pattern is explained in part by the distribution practices of the Council. Print copies are disseminated primarily within the two Canadian Maritime provinces and three American states in the Council's jurisdiction. Readers outside of the Gulf region, if they become aware of GOMC publications, are far more likely to discover the digital publications rather than printed reports, and journal articles are more easily located. The citation pattern also relates to the focus of GOMC's publications on coastal ecosystem and management issues, which may be sufficiently unique to the Gulf of Maine and northwest Atlantic that the publications are of lower relevance to other coastal regions of the world.

Table 2: Citations to GOMC Publications* by Region of the Citing Authors

Region	Online GOMC publications**	Print only GOMC publications
Gulf of Maine Region	157	169
Canada (outside Gulf region)	24	34
USA (outside Gulf region)	58	18
Europe	24	2
Rest of world	14	5
Total	**277**	**228**

* Only "Council Publications" and journal articles noted in Table 1 were used as targets in the citation analysis reported in this paper.

****** Most online publications were also published in print editions

The usage patterns revealed by the citation data are informative in other respects. When citations are charted over time and the types of citing documents are noted (Figure 1), it is apparent that GOMC publications are receiving increasing attention, but many of the citing publications are grey literature rather than peer-reviewed journals and books. In other words, GOMC's grey literature is cited more by other grey literature, and this pattern holds up whether one considers total citations or total citing documents (Figure 2). Very few reports from the early 1990s are available on the web, so citations in grey literature from that period are under-represented in our results.

Figure 1: Total Citations to GOMC Publications, 1991-2006 (n=505)

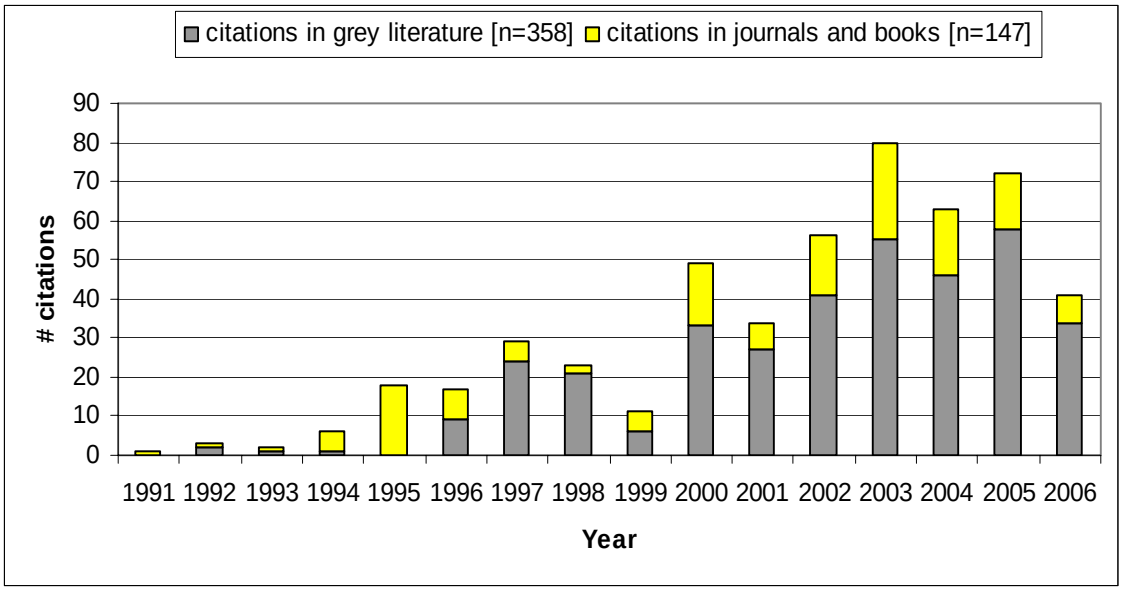
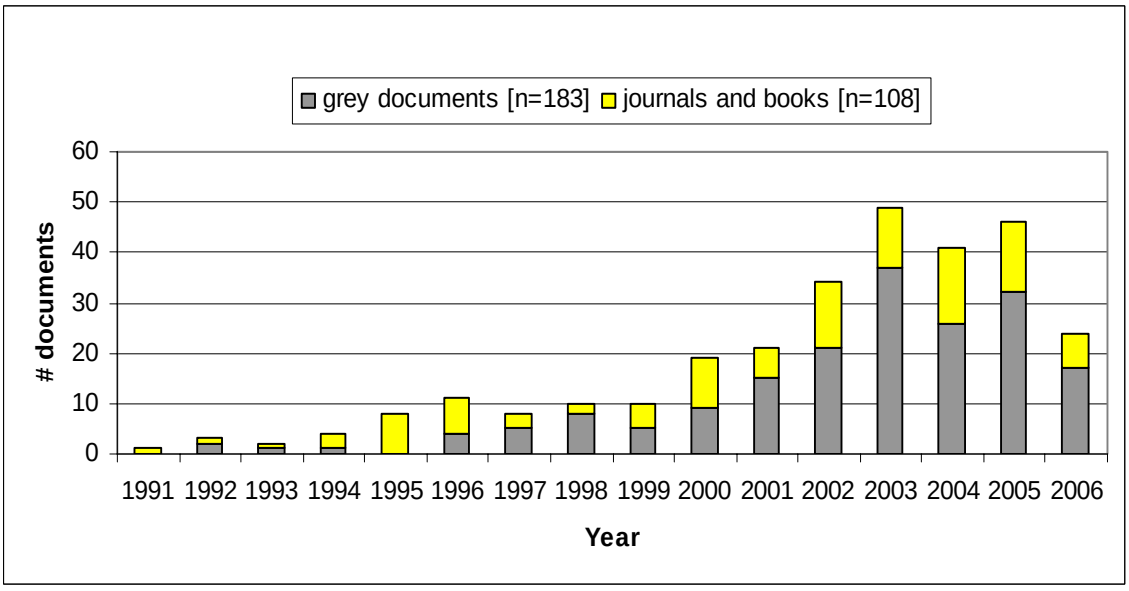


Figure 2: Total Documents Citing GOMC Publications, 1991-2006 (n=291)



The nature of citations to GOMC publications can be probed further in an examination of reports generated by one the Council's significant ongoing initiatives, the Gulfwatch program, noted above. The Council's publications from this program consist of a mix of grey literature and papers published in peer reviewed journals. Detailed data reports are produced periodically, and syntheses of these reports have been published as scientific papers in the leading scientific journals, *Marine Pollution Bulletin* (Chase, et al., 2001) and the *Journal of Shellfish Research* (Jones, et al., 2001). Scientists involved in the Gulfwatch program feel compelled to publish in both genres. The grey literature reports provide a venue for detailed annual documentation of findings, whereas the credibility and accessibility of peer-reviewed periodicals offer heightened exposure for the research undertaken in and major results of this program. The journal article by Chase et al. (2001) is available electronically, and is much more likely than the reports to be cited outside the region. It is cited in 16 articles from Europe, often providing a comparison for the findings of local monitoring programs (e.g., Carro, Cobas & Maneiro, 2006; Green & Knutzen, 2003).

Figure 3: Citations in Journal Articles to Gulfwatch Publications (n = 43)

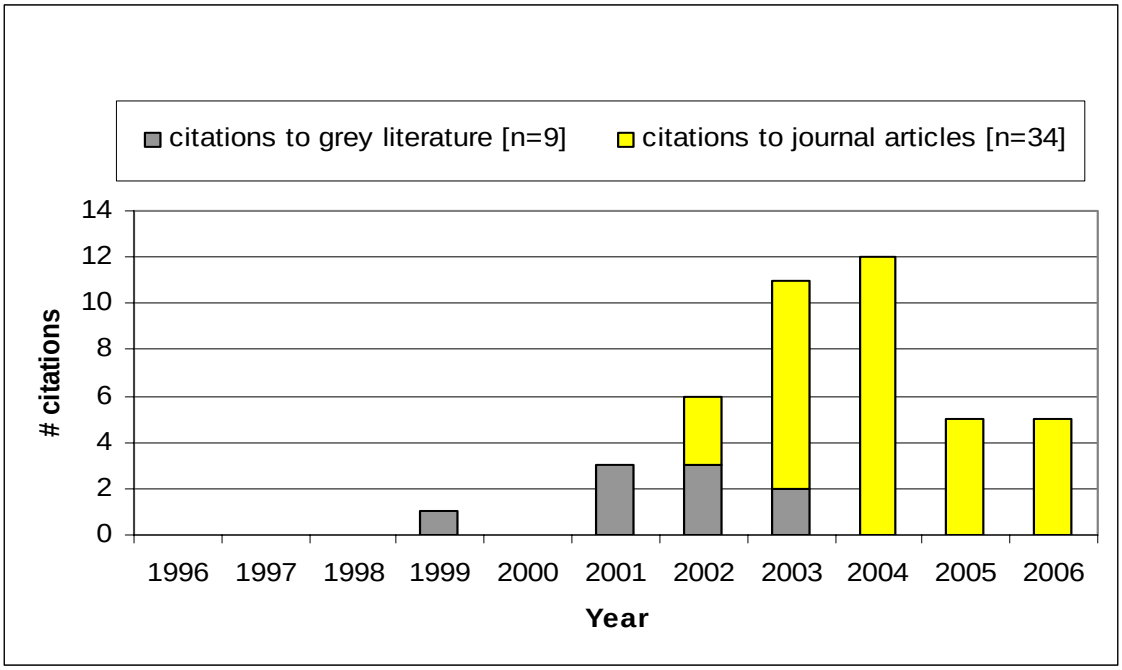
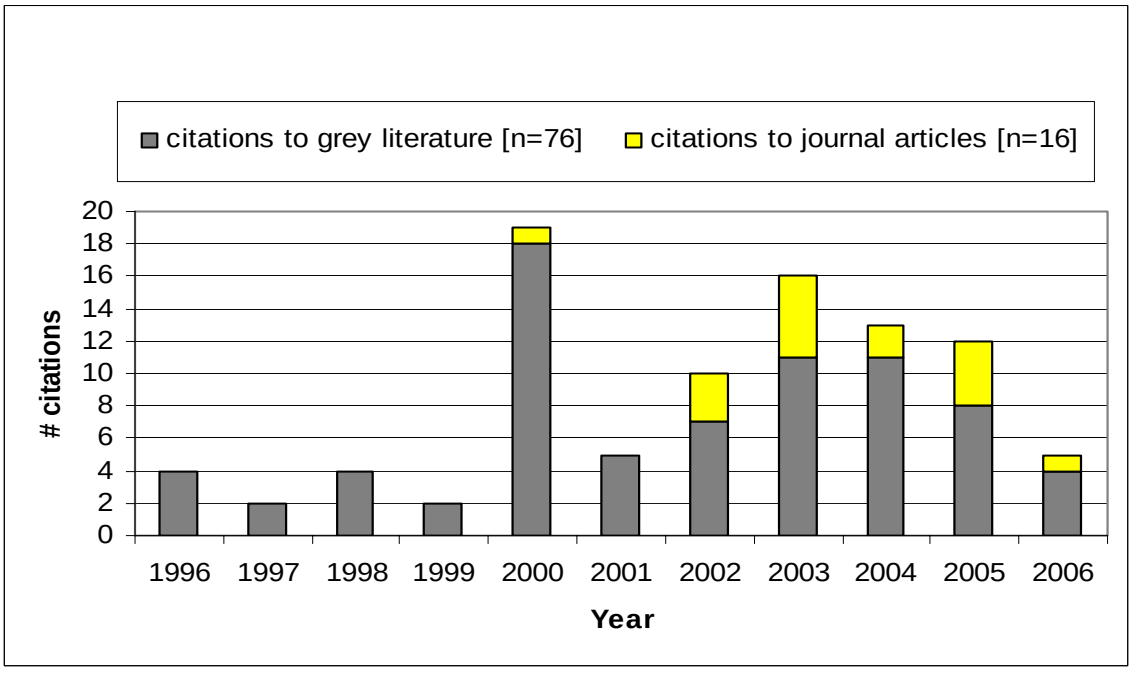
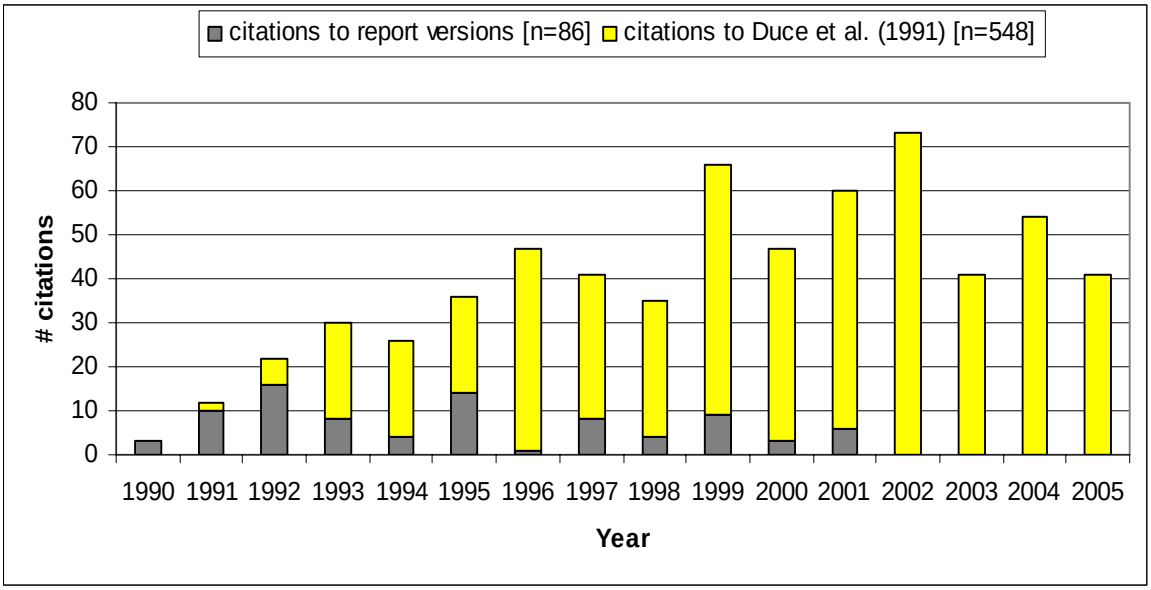


Figure 4: Citations in Grey Literature to Gulfwatch Publications (n = 92)



Figures 3 & 4 show clearly a bias in citation patterns**.** Papers in scientific journals are much more likely to cite the Gulfwatch journal articles than the grey literature on the same topic, which may provide more detailed data and analysis (Figure 3). Similarly, authors of grey literature are more likely to cite other grey literature than journal publications (Figure 4). This citation bias was also found in the study of citations to GESAMP publications (Figure 5). In the GESAMP case, authors of scientific papers were far more likely to cite the version of *The Atmospheric Input of Trace Species to the World Ocean* published by Duce et al. as in a scientific paper (Duce et al., 1991), than the original report (GESAMP, 1989). Since GESAMP reports are rigorously peer reviewed, the citation bias is not necessarily related to the perceived quality of a journal article versus a technical report.

Figure 5: Citations to Versions of *The Atmospheric Input of Trace Species to the World Ocean**



***** Adapted from MacDonald, Cordes & Wells (2004). Citations to the report versions (GESAMP, 1989) were extracted from *Web of Science* to the end of 2001 only. Citations to Duce et al. (1991) were extracted from *Web of Science*, *Scopus,* and *SciFinder Scholar*.

The citation patterns uncovered in both the GOMC and GESAMP examples may be attributed to limited distribution of grey literature versus scientific periodicals, variations resulting from searching tendencies of researchers (i.e., the tendency to search databases of scientific periodicals rather than grey literature sources), and/or perceptions of the quality of grey literature versus peer-reviewed journals on the part of researchers and journal editors. In addition, citations serve to publicize the cited works. Cordes (2004) examined citations to two GESAMP reports that were each published in three print versions: in two report series and as a journal article or a book. In both instances, the version cited by the authors of the report, in articles they wrote, was the most highly cited version overall. Other citing authors chose the version recommended to them in earlier citing papers.

Tracking usage of grey literature in public policy settings is a complex undertaking due largely to the policy-making process, which limits following connections (i.e., citations) within documentary evidence. Much of the policy documentation is internal to units of governments, is never published formally, and hence is not covered by citation databases. Nonetheless, citation evidence can uncover references to grey literature in policy-making settings. Given the mandate of GOMC, its publications are often relevant for policy-making at municipal, state or provincial, or federal levels, and citations confirm use at all three levels as well as in international contexts. For example, at the federal level, Peter Shelley (Vice-President of the Conservation Law Foundation) cited a GOMC publication when he testified to the U.S. Commission on Ocean Policy on July 24, 2002 (Shelley, 2002). GOMC publications are noted within state and provincial policy documentation, such as the watershed assessment report for the Merrimack River in Massachusetts (Dunn, 2001). While international interest is limited, citations show that GOMC publications have been used in marine policy areas in Australia (Baker, 2000), and as background to decisions of the World Court related to the UN Convention on the Law of the Sea (Kwiatkowska, 2002). However, most public policy references to GOMC publications are regionally based, at municipal, state/provincial and federal government jurisdictions.

Since GOMC increasingly is publishing digital versions of documents, statistics captured from traffic on the Council's website are a further indicator of usage of grey literature. Data over a 17-month period from January 2005 through May 2006, in terms of "page views per month" (Figure 6) and "user sessions per month" (Figure 7), show a steady rise in website traffic (P.H. Taylor & J. Cradock, personal communication, July 15, 2006). These statistics emphasize the growing importance of web presence for organizations like GOMC. A further indicator of the significance of the web lies in interconnections of sites on the web. Tracing those interconnections via search engines in any definitive manner is problematic due to indexing and search engine vagaries, but a broad brush perspective about internet links can be seen in the number of links to the GOMC website in comparison to the GESAMP site. On December 1, 2006, a search for links using the Yahoo search engine ("linkdomain:gulfofmaine.org -inurl:gulfofmaine.org") located 1000 links to the GOMC website from other websites, and 486 to the GESAMP site ("linkdomain:gesamp.imo.org -inurl:gesamp.imo.org"). The number of links reflects the more extensive and sophisticated GOMC website, and indicates that GOMC publications may be receiving greater interest and use than GESAMP's.

Figure 6: Page Views per Month (www.gulofmain.org) January 2005 – May 2006
(credit P. Taylor and J. Cradock)

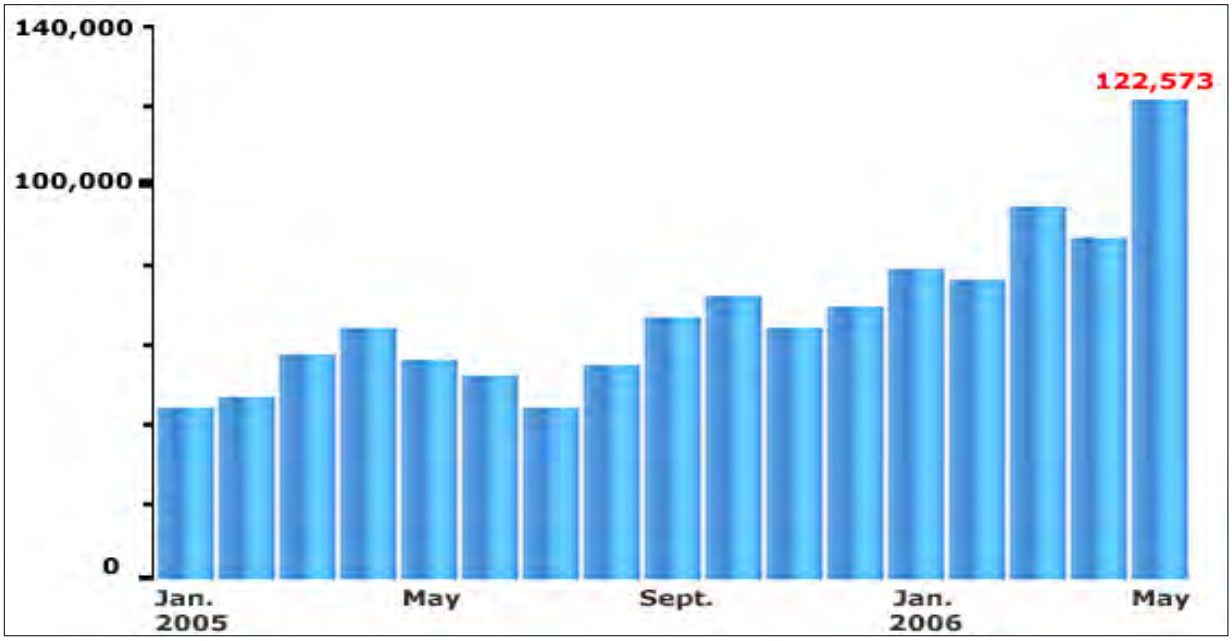
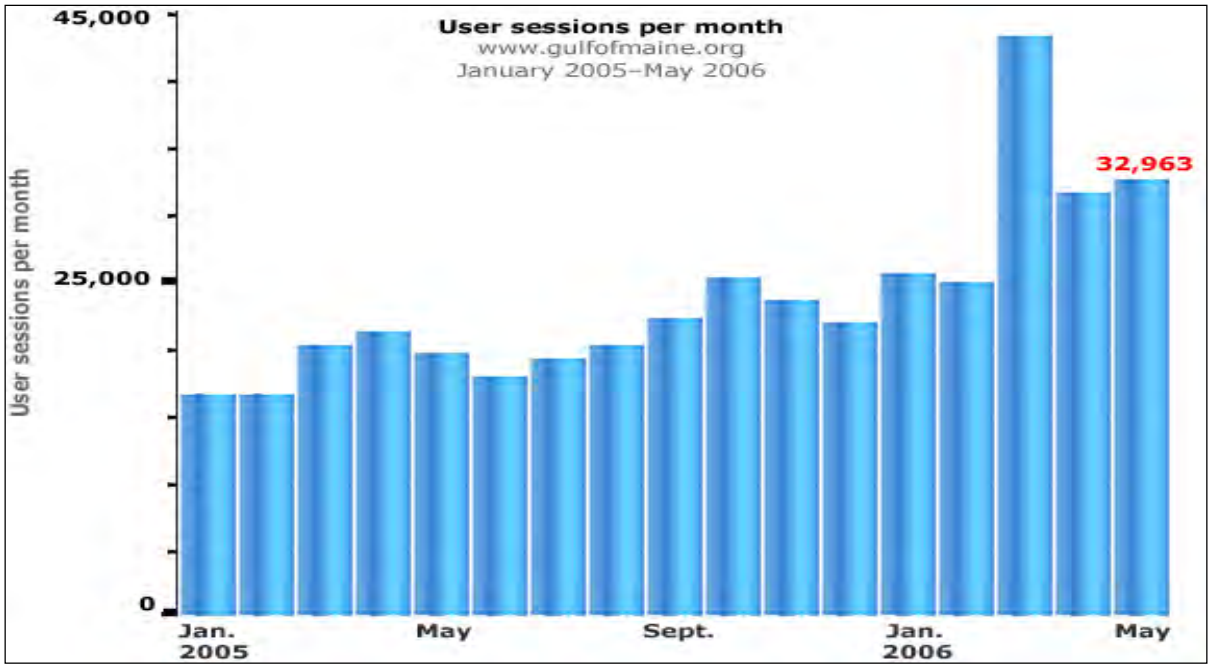


Figure 7: User Sessions per Month (www.gulfofmain.org) January 2005 – May 2006
(credit P. Taylor and J. Cradock)



Evolving citation tools and searching services offer increasing means of locating data that indicate usage patterns. Such tools also promote use of grey literature, particularly when the literature is accessible via the web.

Conclusions

In 2004, Peter Taylor, a science translation writer employed by the Gulf of Maine Council, wrote: "A core constraint for understanding and managing the oceans on a regional scale has been information: collection of data for research and monitoring; data sharing; integration and analysis; and regional exchange of findings and management solutions. These challenges are magnified by the geographic size and ecological complexity of a system like the Gulf of Maine" (Taylor, 2004, p.1). Taylor's assessment, which places

information at the centre of environmental management issues, brings the questions raised at the beginning of this paper back into focus. In recognition of the importance of information diffusion and use in environmental management decision making, the following points can be made:

1) Grey literature is the mainstay of the GOMC's significant publication initiatives, which now number more than 300 titles. In this regard, GOMC and GESAMP are very similar. Both marine environmental intergovernmental organizations publish mostly grey literature.
2) In contrast to GESAMP, which primarily produces rigorously refereed reports (but still by definition grey literature), GOMC generates a variety of information products, not all of which are refereed.
3) For both GOMC and GESAMP, dissemination of print publications has been uneven. GOMC has been more aware than GESAMP of the value of effective information dissemination, but neither organization has achieved consistency in practice.
4) With continuing development of its website (which is itself a rich information source), GOMC may be overcoming dissemination problems. Further study of web access traffic and web links will confirm whether usage is wider spread than evidence drawn from citation data currently shows. In contrast to GOMC's increasing web-based initiatives, GESAMP and the UN agency secretariat which supports its publication programme have devoted very limited resources to website development and maintenance.
5) Evidence of usage of GOMC publications is manifested in the assembled citation data. Citation patterns for GOMC and GESAMP illustrate similarities of publication use. Characteristics in citations to GOMC publications highlighted a dichotomy regarding the source of citations, which suggests that grey literature may be overlooked or its use discouraged in peer-reviewed scientific literature. Why this dichotomy exists warrants further study.
6) While evidence that GOMC publications are used in public-policy settings was uncovered, the evidence is not strong. However, even if more comprehensive searches for citations were undertaken, this investigative technique is unlikely to yield significantly greater understanding of usage patterns. Citation analysis has been informative, but it has limitations for tracking public policy and management documentation in contrast to research literature (see, for example, Bertrand, and Côté, 2006).
7) To complement our current findings, other methods of determining usage in policy decision making contexts are needed. GOMC was set up for direct transfer of scientific information to public sector managers. Senior policy personnel participate in the Council meetings and are provided with a sizeable volume of documentation. Tracking the life of information contained in a GOMC publication through the Council meeting documentation to other policy documents would provide additional insights on the impact of the Council's work.

Since finding solutions to the environmental problems that are threatening planetary health will rest in part on effective transfer of scientific findings and knowledge into public policy, grappling with the challenges that grey literature poses is vitally important. Will greater dependence on digital publications accessible on the web provide the solutions? The complexity of the phenomenon of information creation, distribution and incorporation into knowledge and action implies that the answer to this question will be elusive, and no single answer will be sufficient. That the problem deserves an answer is beyond question and will form the focus of future studies.

References

Acreman, M. (2005). Linking science and decision-making: Features and experience from environmental river flow setting. *Environmental Modelling & Software, 20*, 99-109.

Baker, J. K. (2000). *Guide to marine protected areas*. [S.l.]: South Australia Department for Environment and Heritage. Retrieved December 12, 2006, from http://www.environment.sa.gov.au/coasts/pdfs/mpa1.pdf

BBC News. (2006a, October 30). Expert reaction to Stern review. Retrieved December 12, 2006, from http://news.bbc.co.uk/1/hi/business/6098612.stm]

BBC News. (2006b, October 31). Climate change fight "can't wait." Retrieved December 12, 2006, from http://news.bbc.co.uk/2/hi/business/6096084.stm

Bertrand, F. & Côté, G. (2006). 25 years of Canadian environmental research. A scientometric analysis (1980-2004). Ottawa: Science-Metrix. Retrieved December 12, 2006, from http://www.science-metrix.com/pdf/SM_2006_001_EC_Scientometrics_Environment_Full_Report.pdf

Carro, N., Cobas, J., & Maneiro, J. (2006). Distribution of aliphatic compounds in bivalve mollusks from Galicia after the Prestige oil spill: Spatial and temporal trends. *Environmental Research, 100*, 339-348.

CBC News. (2006, October 30). British government report: Global warming will devastate world economy. Retrieved December 12, 2006, from http://www.cbc.ca/cp/science/061030/g103022.html

Chase, M. E., Jones, S. H., Hennigar, P., Sowles, J., Harding, G. C. H., Freeman, K., Wells, P. G., Krahforst, C., Coombs, K., Crawford, R., Pederson, J., & Taylor, D. (2001). Gulfwatch: Monitoring spatial and temporal patterns of trace metal and organic contaminants in the Gulf of Maine (1991-1997) with the blue mussel, Mytilis edulis L. *Marine Pollution Bulletin, 42*, 491-505.

Cordes, R. E. (2004). Is grey literature every 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,* 45-65.

Cordes, R. E., MacDonald, B. H. & Wells, P. G. (2006). *Publications of the Gulf of Maine Council on the Marine Environment and their use*. Unpubl. rept. for the Gulf of Maine Council on the Marine Environment. Halifax: Dalhousie University.

Doern, G. B. & Conway, T. (1994). *The greening of Canada. Federal institutions and decisions*. Toronto: University of Toronto Press.

Duce, R. A., Liss, P. S., Merrill, J. T., Atlas, E. L., Buat-Menard, P., Hicks, B. B., et al. (1991). The atmospheric input of trace species to the world ocean. *Global Biogeochemical Cycles 5*(3): 193-259.

Dunn, W. J. (2001). *Merrimack River: A comprehensive watershed assessment report*. [S.l.: Massachusetts Executive Office of Environmental Affairs. Retrieved December 12, 2006, from http://www.mass.gov/envir/water/publications/assessment_reports/Merrimack_Assessment_Report_2000.pdf

Elsevier. (2006). Using Scopus for bibliometric analysis: A practical guide. *Library Connect, 9*.

Francis, T. B., Whittaker, K.A., Shandas, V., Mills, A.V., & Graybill, J.K. (2005). Incorporating science into the environmental policy process: A case study from Washington State. *Ecology and Society, 10*(1):35, Retrieved December 12, 2006, from http://www.ecologyandsociety.org/vol10/iss1/art35/

GESAMP. (1989). *The atmospheric input of trace species to the world ocean*. Reports & Studies, No. 38. (Reprinted 1990, UNEP Regional Seas Reports & Studies, No. 119).

Green, N.W., & Knutzen, J. (2003). Organohalogens and metals in marine fish and mussels and some relationships to biological variables at reference localities in Norway. *Marine Pollution Bulletin, 46*, 362-374.

Gulf of Maine Council on the Marine Environment. (1993). *Annual report, 1992 / 1993.* [S.l.]: Author.

Gulf of Maine Council on the Marine Environment. (1996). *Work in progress: Five-year report of the Gulf of Maine Council on the Marine Environment, 1990-1995*. [S.l.]: Maine State Planning Office / 1995-96 Secretariat Jurisdiction, Gulf of Maine Council on the Marine Environment.

Jones, S. H., Chase, M., Sowles, J., Hennigar, P., Landry, N., Wells, P. G., Harding, G. C. H., Krahforst, C., & Brun, G. L. (2001). Monitoring for toxic contaminants in Mytilus edulis from New Hampshire and the Gulf of Maine. *Journal of Shellfish Research, 20,* 1203-1214.

Kwiatkowska, B. (2002). *Decisions of the World Court relevant to the UN Convention on the Law of the Sea: A reference guide*. The Hague: Kluwer Law International.

MacDonald, B. H., Cordes, R. E., & Wells, P. G. (2004). Grey literature in the life of GESAMP, an international marine scientific advisory body. *Publishing Research Quarterly, 20*, 26-41.

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being. Synthesis*. Washington, DC: Island Press.

O'Dell, C., Dallman, D., Vesely, M. & Vigen, J. (2003). 50 years of experience in making grey literature available. Matching the expectations of the particle physics community. GL5: Fifth International Greynet Conference, Amsterdam, December 2003. Amsterdam: TextRelease.

Pesch, G. G. & Wells, P. G. (Eds.) (2004). *Tides of change across the Gulf. An environmental report on the Gulf of Maine and Bay of Fundy*. Augusta, Maine: Gulf of Maine Council on the Marine Environment. Retrieved December 12, 2006, from http://www.gulfofmaine.org/council/publications/tidesofchangeacrossthegulf.pdf

Roux, D. J., Rogers, K. H., Biggs, H. C., Ashton, P. J. & Sergeant, A. (1969). Bridging the science-management divide: Moving from unidirectional knowledge transfer to knowledge interfacing and sharing. *Ecology and Society, 11*: 4, Retrieved December 12, 2006, from http://www.ecologyandsociety.org/vol11/iss1/art4/

Schrecker, T. (2001). Using science in environmental policy: Can Canada do better? In Parson, E.A. (Ed.). *Governing the environment: Persistent challenges, uncertain innovations.* (pp. 30-70). Toronto: University of Toronto Press.

Shelley, P. (2002). *Looking forward: Developing a federal strategy, vision and operational capacity for managing the Gulf of Maine*. (Testimony before the U.S. Commission on Ocean Policy). Retrieved December 12, 2006, from http://www.oceancommission.gov/meetings/jul23_24_02/shelley_testimony.pdf

Springer, A. L. (2002). *North-American transjurisdictional cooperation: The Gulf of Maine Council on the Marine Environment.* Canadian-American Public Policy, An Occasional Paper, no. 50.

Stern, N. (2006). *Stern review on the economics of climate change.* London: HM Treasury. Retrieved December 12, 2006, from http://www.hm-treasury.gov.uk /independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm

Taylor, P. H. (2004). Data sharing and management. *Gulf of Maine Times, 8*(2). Retrieved December 12, 2006, from www.gulfofmaine.org/times/spring2004/science_insights.html and http://www.gulfofmaine.org/council/publications/scienceinsights/datasharing.pdf

Tyrrell, M. (2005). *Gulf of Maine marine habitat primer*. [S.l.]: Gulf of Maine Council on the Marine Environment. Retrieved December 12, 2006, from http://www.gulfofmaine.org/habitatprimer.

Van Duesen, K. & Johnson Hayden, A. C. (1989). *The Gulf of Maine: Sustaining our common heritage*. Augusta, ME: Gulf of Maine Council on the Marine Environment.

Wake, C., Burakowski, L., McKenzie, K. & Huntington, T. (with Burtis, B.). (2006). *Cross border indicators of climate change over the past century: Northeastern United States and Canadian maritime region.* [S.l.]: Gulf of Maine Council on the Marine Environment. Retrieved December 12, 2006, from http://www.gulfofmaine.org/council/publications/cross-border-indicators-of-climate-change.pdf

Awareness and empowerment in document production and distribution as a "must" for open access: experiences from the "Nancy style" to guarantee quality

Paola De Castro and Sandra Salinetti
Publishing Activities, Istituto Superiore di Sanità, Italy

Marcus Banks
New York University Medical Center, New York, United States

Background

The debate on Grey Literature (GL) has now a very long tradition going from uncertainty and confusion, in the last century, to new certainties and appraisal in the open access era. This implies the general acquisition of the awareness of the importance of GL as a fundamental primary source of information, hence the necessity to empower authors and issuing organization to guarantee quality in both production and diffusion of GL. During the last international conference on GL held in Nancy in 2005, the proposal for the adoption of an *ad hoc* GL production style was discussed and most welcome by the grey community.[1] This led to the creation of the Grey Literature International Steering Committee (GLISC) which issued, in March 2006, the *Guidelines for the production of scientific and technical reports*,[2] informally known as "Nancy style" (freely available from www.glisc.info).

In the creation of the *Guidelines*, suggestions arising from different realities were closely evaluated and it was not always so easy to reach an agreement that would satisfy all requirements. When the first version was released, there was awareness that some missing points should be developed soon, particularly all issues related to digital GL and metadata.

Goal

This paper intends to reflect on the necessity to update the *Guidelines*, one year after the first draft was presented in Nancy, taking into consideration the rapid changes in the information market. We are convinced that the *Guidelines* may represent a useful tool to improve quality of GL, increase the general awareness of its value, and empower authors and issuing organizations to allow a correct autonomous production of documents to be spread and exploited at best through the Internet. For the same reason, the best strategies to promote the adoption of the *Guidelines* will be investigated: involving GL authors and institutions, making translations available in different languages, establishing useful contacts with key persons in the information arena.

Actions developed

With the objective to create major awareness on quality issues in GL creators, users, and policy makers, the following actions were developed:

- analysis of GL in the changing context;
- reflections on the development of the "Nancy style";
- comparison between "Nancy style" and ANSI/NISO Z39.18;
- promotion of the "Nancy style".

Analysis of GL in the changing context

Within the hierarchy of information sources, GL has traditionally occupied a less prestigious status than peer-reviewed papers published in scholarly journals. During the print-only era, this was a valid distinction. Even though GL has always provided access to valuable and unique content, the costs in time and money of obtaining it were formidable. In such a climate, the regularity and reliability of peer-reviewed journals were significant advantages, irrespective of the quality of published papers.

The contemporary ease of electronic distribution has erased a key condition for the traditional hierarchy between grey and white literature. Although this point has not been reached yet, theoretically all scholarly content – be it a data set, a working paper, or a formal paper – could instantly be posted online.

The "open access" movement has focussed on increasing online availability of formal papers, but its general principle is applicable to all scholarly content. Even though distribution of all content has eased dramatically, the firm preference among many scholars for formally peer-reviewed content continues. This is true despite evidence that the consideration of GL enriches many scholarly investigations;[3,4] one of its

[1] De Castro P, Salinetti S. "Uniform Requirements" for grey literature: proposal for the adoption of "Nancy style". *Publishing Research Quarterly* 2006;22(1):12-7.

[2] Grey Literature International Steering Committee. *Guidelines for the production of scientific and technical reports. Version 1.0*. GLISC; 2006. Available at: http://www.glisc.info; last visited Nov. 14, 2006.

[3] Conn VS, Valentine JC, Cooper HM, Rantz MJ. Grey Literature in meta-analyses. *Nursing Research* 2003;52(4):256-61.

most valuable characteristics is that it is more likely to report negative results, and to discuss studies that concluded prematurely.[5] Compared to relying only on published articles, also accessing GL offers a much more comprehensive picture of the state of knowledge for any topic.

Because GL is so useful, raising its profile is imperative. One way that the authors of reports can achieve this is to adhere to well-developed production standards. This was the motivation for the development in Europe of the "Nancy style" at the GL7 Conference in December 2005, and for refinement of this style in 2006. In the USA, the *ANSI/NISO Standard Z39.18-2005 Scientific and Technical Reports – Preparation, Presentation, and Preservation*,[6] serves a similar purpose.

The encouragement of quality production standards is an "internal" advance among GL producers. Current developments in the external information landscape also point to an enhanced public awareness. In October 2006 the Canadian Institutes of Health Research (CIHR) released a "Draft Policy on Access to CIHR-funded Research Outputs", which operates from the premise that "the primary purpose of all research in the public domain is the creation of new knowledge in an environment that embodies the principles of freedom of inquiry and unrestricted dissemination of research results" (http://www.cihr-irsc.gc.ca/e/32326.html). The draft policy calls for peer-reviewed articles based upon CIHR-funded research to be openly available online no longer than six months after publication. This is similar to the currently pending "Federal Research Public Access Act of 2006" in the United States (http://cornyn.senate.gov/doc_archive/05-02-2006_COE06461_xml.pdf). The US measure only applies to peer-reviewed journals. Critically, the Canadian measure *also* requires grant recipients to provide access to "research materials" and "research data". Research materials might include cell lines, nucleic acids, or research tools for evaluation. Research data would usually be an electronic data set, but could also include interview transcripts and survey results.

These research materials and research data are different examples of "grey content", which is the kernel of GL. If the Canadian policy goes into effect, and grey content is exchanged among researchers as a matter of course, this would revolutionize scholarly norms. The Canadian proposal is one example of the movement toward "open data" in science, which seeks to ensure maximum access to several types of data:

- *scientific data* "deemed to belong to the commons" (i.e., the human genome);
- *infrastructral data* such as that provided by geographic information systems;
- *factual data* that is not copyrightable (http://wwmm.ch.cam.ac.uk/blogs/murrayrust/?p=32).

GL producers and users should become strong proponents of the open data movement as well. The Open Data Foundation is one organization working in this area (http://www.opendatafoundation.org/).

A leading proponent of the open data movement is Peter Murray-Rust of the University of Cambridge. Murray-Rust is also one of the developers of the concept of "datuments", which are scientific papers enriched with data elements that can only be accessed and interpreted electronically (http://jodi.tamu.edu/Articles/v05/i01/Murray-Rust/). Datuments integrate traditionally grey and white content into one package, and their proliferation would be strong evidence of the end of the distinction between grey and white literature. As argued at GL7,[7] this distinction would eventually fade with time.

Time will tell whether all this is correct. In the meantime, paying careful attention to the "Nancy style" and ANSI/NISO standard is an excellent way to support the promotion of GL.

Reflections on the development of the "Nancy style"

The difficulties encountered during the creation of the *Guidelines* were analysed to reflect on the selection process required to reach consensus at international level, even within a small group of experts in a specific field. The major points of discussion mainly regarded the electronic GL and adoption strategy:

- **Electronic grey literature**
  The "Nancy style" is mostly paper oriented, because editorial consistency and ethical considerations recommended for traditional documents do apply also to digital publications. Yet, progressively more and more GL is being produced, stored, published and made available electronically and in order to manage relevant GL publications, metadata are required. The importance of metadata, as the natural evolution of library catalogue records, had been already stressed in the first version of the "Nancy style" (when dealing with report structure: Section 4.2 of the *Guidelines*), but no metadata schema was then provided since it was difficult to find a formula that would satisfy all requirements. At present, much GL is catalogued using the Dublin Core Metadata Standard (DC). However – as Keith Jeffery of the UK Council for the Central Laboratory of the Research Councils (CCLRC) pointed out working on the "Nancy style" draft – this standard suffers from several problems: a) it is machine-

[4] Benzies KM, Premji S, Hayden KA, Serrett K. State-of-the-evidence reviews: advantages and challenges of including grey literature. *Worldviews on Evidence-based Nursing* 2006;3(2):55-61.

[5] See ref. 3.

[6] ANSI/NISO. *Scientific and Technical Reports – Preparation, Presentation, and Preservation*. Bethesda, MD: NISO Press; 2005. (Standard Z39.18-2005). Available at: http://www.niso.org/standards/resources/Z39-18-2005.pdf; last visited Nov. 14, 2006.

[7] Banks M. Towards a continuum of scholarship: the eventual collapse of the distinction between grey and non-grey literature. *Publishing Research Quarterly* 2006;22(1): 4-11.

readable but not machine-understandable; b) it does not have a formalised syntax or semantics and therefore is open to ambiguous interpretations. Therefore, he proposed a formalised metadata standard (an umbrella standard, mainly generated from Dublin Core metadata: "Formalised DC" based on the concepts of the CERIF Model (www.eurocris.org/Cerif). Yet, as the traditional cataloguing practice has different rules, similarly different communities may adopt different metadata schema. Nowadays the World Wide Web provides the possibility to search for information across heterogeneous archives/databases/catalogues, but the systems managing different information resources must be "interoperable" (capable to work together), and interoperability requires that the same metadata schema be used. As Stefania Biagioni (of the Italian Istituto di Scienza e Tecnologie dell'Informazione - ISTI, Consiglio Nazionale delle Ricerche) clearly commented, there is much work towards standardization and the Dublin Core Initiative (http://dublincore.org/) is receiving worldwide consensus as it suggests adding a very simple metadata record to any specialized one.

- **Adoption strategy**
  When consensus was to be reached to release the first version of the *Guidelines*, a formal approval was asked to all organizations wishing to officially adopt them. Contrary to expectations, consensus was given only by a small number of institutions as the official adoption was sometimes a difficult step. Yet, support and encouragement did not lack: a less formal apprach in launching the *Guidelines* and getting them adopted was soon granted by all institutions involved in their creation. For example, a large international organization (Organisation for Economic Co-operation and Development - OECD), which took part in the development of the *Guidelines*, expressed concern to officially endorse them (and in fact, it did not), because that would require a great deal of internal debate and discussion with their own members. Suggestions were made to follow a voluntary system backed up by an official recognition of compliance to facilitate the adoption of the *Guidelines*. This would encourage like-minded supporters within an organisation to informally use the *Guidelines* and then gain the official "stamp of approval" to show that they are really following them. Actually, other organizations policies take a voluntary approach in the documents they recommend, such as the Association of Learned and Professional Society Publishers (ALPSP) with more than 230 not-for-profit publishers. As suggested by the OECD, voluntary sign-up is a less demanding step for organisations to take, but the effect is the same – more and more publishers will opt to use them.
  In the next version of the "Nancy style" the adoption strategy may be revised, envisaging a voluntary approach followed by a subsequent official endorsement, once the organisation can show that *Guidelines* are actually followed.

Comparison between "Nancy style" and ANSI/NISO Z39.18

During the working process leading to the production of the *Guidelines*, and also in the period soon after their publication, we came in touch with other standards, guidelines, house styles, etc.
All useful documents were further studied in view of the future updating of the *Guidelines*. In particular, the *ANSI/NISO Standard Z39.18-2005 Scientific and Technical Reports – Preparation, Presentation, and Preservation* (released in 2005) has been considered a valuable source for comparison; therefore, we carefully analysed it and pointed out the main differences listing them under the following headings.

General considerations

The major differences concerning the two documents as a whole regard:

- **Document type**
  They are different in that the "Nancy style" represents *guidelines* – that is general principles agreed upon by a small group of experts, to be followed as an indication or outline of policy or conduct –, while the ANSI/NISO Z39.18 is a proper *standard*, developed by the Standards Committees of the US National Information Standards Organization (NISO), subject to rigorous control and approval process including peer review. This is why also the structure of the two documents is different since the standard may repeat concepts in different sections which may be used separately, while the *Guidelines* are intended as an easy to read document giving the general idea for recommended items. The *Guidelines*, different from standards, do not give full details on format and style.
  Moreover, the "Nancy style" represents *international* guidelines developed by a corporate author (GLISC), which worked on the draft proposed by the Istituto Superiore di Sanità, and signed approval of this best practice on behalf of their respective organizations, while the ANSI/NISO Z39.18 is a *national* standard approved by the American National Standards Institute through a number of Voting Members.

- **Paper *vs* digital document medium**
  The "Nancy style" is mostly paper oriented giving recommendations on report preparation mainly reflecting a traditional paper structure, while the organization pattern of the ANSI/NISO Z39.18 is

user-based more than content-based. The key concepts incorporated in the American standard mainly refer to metadata, persistence of links, interoperability, creation, discovery/retrieval, presentation in digital format (DTD, XML, XSL), maintenance and preservation (original content, software and media); it also contains a metadata schema, which is absent in the *Guidelines*.

- **Annexes**
  All material included in the "Nancy style" is approved by the GLISC, while the ANSI/NISO Z39.18 provides a large amount of additional information (almost half of the pages) that is not part of the Standard (Appendices including selected annotated bibliography, glossary, Dublin Core data elements, etc.).

Content considerations
In general, the "Nancy style" contains technical requirements for a report, but does not include full details (i.e. format, style, etc.); yet, it provides important elements, which are not present or not fully described in the ANSI/NISO Z39.18:

- **Ethical issues**
  An initial section is explicitly devoted to authorship, editorship, peer review, conflicts of interest, privacy and confidentiality.

- **Instructions to authors**
  Producers are strongly recommended to issue instructions to guide authors in the production of a formally correct document containing ethical and editorial issues as well as indications for formats, styles, illustrations, etc.

- **Revision**
  Special attention is given to revision editing as GL is not generally peer reviewed, or produced with editorial support; therefore, it is fundamental that authors be aware of the importance of a careful revision of their texts before diffusion.

- **Reference style**
  The adoption of the "Vancouver style" is recommended and examples and rules are given as a fundamental step for information retrieval.

As regards document structure, it is basically the same in "Nancy style" and ANSI/NISO Z39.18, with minor terminological variations. Yet, the American standard explicitly gives indication on:

- Report Documentation Page (since it is used by some agencies within the federal government, and also some sample pages are given).
- Distribution list.
- Glossary (although not part of the Standard).
- Executive abstract.

Technical recommendations
Since the "Nancy style" represents guidelines and not a standard, all technical considerations are limited to the essential, while the ANSI/NISO Z39.18 gives indications (all absent in the "Nancy style") on:

- **Print-specific / non-print-specific recommendations**
  The Section 6 "Presentation and display" describes standard methods for ensuring consistency in presentation including designing visual and tabular matter, formatting, etc. and makes a distinction between rules applicable to all reports regardless of mode of publication (paper or digital) and rules applicable to reports published in paper form only.

- **Format**
  Specific information is provided on fonts, line length, margins, page numbering, style, units and numbers, formulas and equations, paper (format and type), printing equipment, ink.

The ANSI/NISO Z39.18 also includes specifications on index entries and errata, which are not present in the "Nancy style".

Promotion of the "Nancy style"
When the "Nancy style" was released, we tried to promote it at best through all possible channels: announcements circulated in different forums; a logo for GLISC was created and its site was developed

(now appearing high ranked in Google and also generating many links); translations in French (by the Institut de l'Information Scientifique et Technique) and Italian (by the Istituto Superiore di Sanità) were made available; and training courses were held using the *Guidelines* as a reference tool.
All these initiatives permitted to spread information about GL in environments where very little was known on this emerging communication channel or it was completely disregarded. Most often, people realized that they used and produced GL without being aware of all the implications laying behind it.
In particular, within the world of scientific editors, much interest in grey matters was shown by professional editors, who were generally unaware of all concerns relating to GL. Following such interest, they even asked to write a chapter on GL for the *Science Editors Handbook.*[8] This book is edited by the European Association of Science Editors (EASE), the major professional association of editors in Europe, defining itself as "an internationally oriented community of individuals from diverse backgrounds, linguistic traditions and professional experience who share an interest in science communication and editing". This chapter was most welcome by all members of the Association and they also asked to spread more information on GL through the journal *European Science Editing.*[9]
The possibility was also envisaged that the new born "Nancy style", like "Vancouver style", might develop into a *de facto* standard, representing uniform requirements for the production of technical reports produced by an international group (GLISC), in the effort to combat ignorance of the best editorial practices and permit integration, cooperation and standardization. Following such enthusiasm, a proposal was made that the direct promotion of the "Nancy style" by EASE might be included in the wider programme of penetration into the world of European science editing, even if GL is traditionally considered as a *sui generis* editorial product.
In the scientific arena, where career is guaranteed by publications in high impact journals, GL plays an important role not only as a vehicle of information, but also as a first step towards autonomous productions of editorially sound documents. Training inexperienced authors to allow a correct production of grey material may represent the best way to reach publications of higher levels. Tutoring in scientific writing and data presentation, in fact, also helps improving research methodology. The experience, gained in years of writing courses in the biomedical field, gives evidence that when scientists understand the reasons why editorial rules and standards must be applied (for both grey or white production), their publication output will rapidly improve as well as the challenges to have an article accepted by mainstream journals.
To date, most of the promotional efforts for the "Nancy style" have occurred in Europe. At the GL8 Conference in New Orleans, authors of this paper will discuss how to increase awareness of the "Nancy style" in North America.

Proposals for updating the "Nancy style"

From the analysis of the above described actions (Analysis of GL in the changing context; Reflections on the development of the "Nancy style"; Comparison between "Nancy style" and ANSI/NISO Z39.18; Promotion of the "Nancy style"), we propose the following suggestions to update the *Guidelines*, which we recognise to be the best tool to promote awareness and empowerment of GL creators:

- **Adding an Appendix on metadata**
  As regards electronic GL, the adding up of an Appendix containing metadata based on Dublin Core is proposed. The Appendix will give a very simple structure representing a general model to follow without going into details. An example for descriptive (bibliographic) metadata regarding an entire issue and a contribution is given in Table 1. This possible form (based on Biagioni's suggestions) will be discussed at GL8. These metadata should be integrated with administrative (rights and software) and structural metadata (hierarchical levels, information used to display and navigate digital resources, etc.), according to the ANSI/NISO Z39.18 structure.

Table 1. Descriptive metadata schema suggested for an entire issue and a contribution (to be discussed at GL8)

Type of document	Descriptive metadata based on Dublin core
Entire issue	dc:title dc:date.issued dc:date.issued.human dc:relation.ispartof.publication.desc dc:relation.ispartof.publication dc:relation.ispartof.volume

[8] De Castro P, Salinetti S. Writing and issuing Grey literature. Chapter 1-3.6. In: *Science Editors Handbook*. European Association of Science Editors; 2006.
[9] De Castro P. Scientists produce and use grey literature, but are they aware of the implications of doing so? *European Science Editing* 2006;32(4):95-7.

	dc:relation.ispartof.year dc:publisher dc:contributor dc:source dc:format
Contribution in an issue	dc:title dc:creator dc:identifier dc:description.fulltext dc:description.abstract dc:format.extent.pagination dc:date.created dc:date.human dc:type dc:relation.ispartof.publication.desc dc:relation.ispartof.publication dc:source dc:relation.ispartof.volume dc:relation.ispartof.issue dc:relation.ispartof.year dc:subject dc:subject.msc dc:subject.jel dc:publisher dc:contributor dc:language dc:relation.ispartof.publication dc:format.extent.pagination

- **Creating a Subject index**
  A Subject index will be very useful to facilitate retrieval of specific items although the online availability permits an easy navigation through the text.

- **Providing more technical advice on digital format**
  Although the philosophy guiding the production of paper documents is unchanged in digital ones, the suggestions given by the ANSI/NISO Z39.18 in Section 6 “Presentation and display” represent a good example to follow as a distinction is made between rules applicable to all reports regardless of mode of publication (e.g., paper, CD-ROM, or Web) and rules applicable to reports published in paper form. A simple table will be added for easy readability including mandatory or optional elements with print-specific and non-print-specific indications.

- **Facilitating reference**
  Since GL is often incorrectly cited, it should be useful to give an indication on how to cite the document using the phrase “To be cited as” followed by the correct citation.

Final considerations

As promoters of the *Guidelines* and members of the GLISC, the experience of working with national and international realities permitted us to reflect more closely on the importance of creating a useful reference document applicable to different contexts.

The *Guidelines* are to be considered as a suggested model rather than a model in itself; they represent a basic step to improve quality in the different stages of GL production in view of its wider electronic circulation. The proposals for their updating will make them more effective, although a regular revision is required to keep pace with the changing ITC scenarios and information policies.

Traditionally editorial rules and ethical considerations were disregarded in the production of GL with negative implications on its quality. The members of the GL community should promote their diffusion mainly within GL issuing organizations that are less aware of existing standards regulating GL production and distribution.

Present and past experiences in GL management and research. A questionnaire survey on Italian participants to GL events.

Rosa Di Cesare, Roberta Ruggieri and Loredana Cerbara
CNR – Istituto di Ricerche sulla Popolazione e le Politiche Sociali, Italy

1. Introduction

The widespread acceptance of Grey Literature throughout Italy is due largely to work undertaken by the Italian Libraries Association (AIB). In 1985, the Association set up a work group to analyse the sphere of grey literature. In 1987, the Association published a monographic issue of its in house magazine dedicated entirely to GL. The issue contained fundamental contributions from Italian library and document bank managers, who, we believe, are destined to become the leading national GL experts. The publication coincided with the Italian participation in System of Information for Grey Literature in Europe (SIGLE) and subsequent establishment of the SIGLE national reference centre at the central CNR library (Di Cesare, Lazzari 2000). Taken together, these initiatives constitute a key focus for the growth of Italian GL experts.
Twenty years on from the initial publication, with the latest redefinition of the role and responsibilities of SIGLE and prior to the eighth international GL conference, it is natural to attempt to gauge past and current Italian contributions in GL at national and international conferences.
It was decided to seek the opinions, of those most qualified to speak from experience and position, on the difficulties and prospects of this sector, which in Italy has undergone fluctuations in fortune.
Above all the enquiry sought to establish how the participating experts view their experience in the study and handling of GL, and what developments they foresee in the sector.
Our qualitative survey aims to provide a profile of Italian GL experts and to gain knowledge useful for future Italian GL initiatives.
In more precise terms, the survey seeks answers to the following interrogatives:

- Quantification and classification of Italian GL experts;
- Quantification of the time they dedicate (or have dedicated) to research on and handling of GL;
- Manner of entry into the field and underlying motivation;
- Self-evaluation of work experience.

As is the case for all qualitative surveys, we did not seek specific indicators to define the appropriacy of GL management policy, nor to measure the efficacy of such policy. Our aim was rather to acquire subjective data from Italian experts and view this data in organisational and cultural contexts.

2. Methods and sample

The survey was carried out by means of a questionnaire sent to all Italian authors who had participated in national and international GL conferences.
The first phase was to analyse authors' contributions at national (Alberani, De Castro, 1993; 1996;1999) and international conferences (GL1 Conference proceedings; GL2, GL3; GL4;GL5; GL6; GL7) between 1992 and 2005. Contributions to the original 1987 monographic issue of the AIB Bollettino were included on account of its historical and pioneering value.
From the above proceedings it was possible to extract names of Italian participants. These number 129 and account for an overall 198 presences at various meetings.
We define participants as those who have signed at least one paper presented at a conference or who were present as moderators. The latter is a very limited category (5), with participation limited to national conferences.

The survey makes use of a semi-structured anonymous questionnaire. It includes 21 open and closed question items along with a small number of scaled response questions. Approximately 15 minutes are required for the questionnaire to be completed.
The questions follow a funnel sequence, from generic to specific, focusing on the field. In this way it is possible to record respondents' opinions and value judgements on their professional experience. These questions were followed by a further section containing structural questions. The final phase was to build a database to load and elaborate the data from the returned questionnaires.

The questionnaire was sent out on September 15^{th,} with deadline for return set at the end of October. Initially there was a healthy number of completed returns, thereafter this number declined significantly. This prompted us to issue a reminder, requesting participants to respond even if some questions were irrelevant in their specific case. The need for such a step had been foreseen and enabled the accumulation of a more than satisfactory total of returns.

Overall there were 129 Italian participants. Given that the survey contemplates a lengthy timescale, it was not sent out to 21 of these, who, in the meantime, had retired or moved to different bodies. It proved

impossible to find them despite the collaboration of their former GL colleagues and attempts to trace them via Google and telephone directories.

This meant that the total number of effective participants was 108. In the majority of cases the questionnaire was e-mailed to participants' work or home addresses. In some cases the questionnaire was sent by snail mail or fax. 59 participants, equalling 54.6% (See Tab. 2), provided correctly completed returns. Three declined to participate on personal grounds.

Tables 1 and 2 illustrate respectively the distribution of participants and respondents per number of papers presented.

Table 1. *Distribution of the (129) Italian GL authors by number of papers*

Number of papers	*Number of authors*	*%*
1	95	73.6
2 - 5	30	23.2
> 5	4	3.1
Total	*129*	*100.0*

Table 2. *Distribution of the (59) respondents by number of papers*

Number of papers	*Number of authors*	*%*
1	38	64.4
2 - 5	17	28.8
> 5	4	6.8
Total	*59*	*100.0*

Tables 1 and 2 show that there is a concentration of respondents in the second and third bands. The two distributions are similar in that they highlight the most numerous figures in the central and final bands (23.2 and 3.1 against 28.8 and 6.8). This indicates that the sample has a denser proportion of experienced respondents than the survey universe. In other words, there is a prevalence of authors of more than one paper presented at the various meetings taken into consideration.
This phenomenon is known as respondent self-selection and is typical of postal surveys such as ours. In the phenomenon the greater the respondents' motivation and involvement, the greater the adherence. In this case, to be precise a qualitative survey, this distortion is not an invalidating factor, but quite the contrary (the sample is not proportionate to the data universe). This is because the data obtained has been provided by the leading experts in the field (from 2 – 5 papers and more than 5 papers) and is therefore far more valuable than that provided by lesser players, who have only presented a single paper.

The above information is completed by Table 3 that illustrates survey coverage in terms of number of papers. It clearly emerges that authors of a single paper are a smaller proportion of respondents. Although we could have given a proportional weighting to this data against the survey universe, we chose not to out of coherence to the qualitative orientations of the survey. It was not our intention to measure a phenomenon but to acquire knowledge on the Italian community of GL researchers.

Table 3. *Coverage of the sample with respect to the number of papers Distribution of the (59) respondents by number of papers*

	Authors		Respondents		Coverage
	n.	%	n.	%	%
Number of papers					
1	95	73.6	38	64.4	40.0
2 - 5	30	23.2	17	28.8	56.6
> 5	4	3.1	4	6.8	100.0

3. Findings

Survey findings begin with definition of a social profile of respondents. This is followed by results of analysis on the first set of questions on time respondents spend on GL research and management. Thereafter we give results for reasons for initial involment into the field and underlying motivation, combined with results of respondents opinions on certain specific issues. The analysis is completed by the results for the final set of questions on Self-evaluation of work experience.
The results are satisfactory both in terms of the number of questionnaires returned (54.6%) and considering the quality of the returns. In most cases the questionnaire was fully and accurately completed. Many respondents went beyond a concise response, giving articulated points of view and taking longer than the requested time.
Accordingly, the 59 respondents provide significant data and confirm the findings of the previous section. Their responses gain further significance given their high educational achievements, illustrated below.

Number and nature of characteristics applying to Italian GL experts

3.1 Social profile

There is a prevalence of women of relatively advanced average age, with a high level of educational qualification and specialisation. (See Tab. 4).
Not only do almost all respondents have a degree (93.2%), as many as 45.8% have a further specialisation, either a Masters degree or a PhD. More than half (54.2%) have attended LG training courses or seminars and no fewer than 47.5% have held such courses or seminars.

Table 4. **Profile of respondents by gender, age and, education (n= 59)**

Gender	**%**
M	23.7
F	76.3
Age	
30-39	10.2
40-49	23.7
50-59	50.8
> 60	15.3
Education	
Master, PhD	45.8
Degree	47.4
High school	6.8

The degrees held are predominantly in the Humanities, but there is also a good proportion of Science degrees (31.0%). This is reflected in the relatively high positions held. Nearly 30% have managerial posts in libraries or documentation centres, while 20% are researchers. Just over half work for public research organisations, while nearly 30% are civil servants. The former category are more frequently found at international meetings, the latter at domestic conferences. (See Tab. 5)

Table 5. **Profile of respondents by sector and, position (n= 59)**

Sector	**%**
University	3.4
Public research institute	52.5
Public administration	30.5
Other	13.6
Position	
University professor	1.7
Library/Documentation centre director	28.8
Researcher	20.3
Librarian	25.4
Technician	8.6
Other	15.2

3.2 Breakdown of involvement and motivation

More than half the respondents are currently working both on research and GL management. Researchers outnumber those involved in GL management. Entry to the sector was determined by in house needs, mainly the in house output of grey literature, but also by personal interest. Respondents believe that there is a link between GL and open access and that this link constitutes a path for the evolution of GL management. They feel their experience is stimulating and enriching and they would choose the GL sector again.

Time dedicated (or that has been dedicated) to research on and management of GL

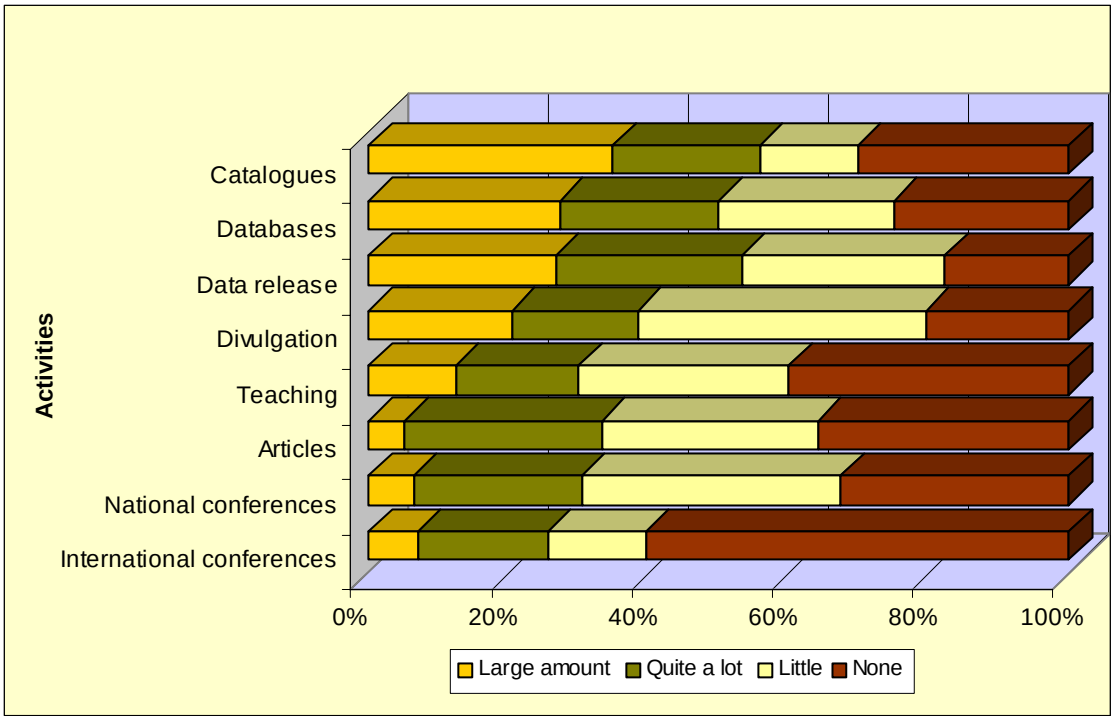
Regarding the time dedicated to research on and management of GL, it should be noted that those spending most time are researchers (See Tab. 6). This element also emerges in responses to the question, "Would you choose to work in GL again?", at the end of the questionnaire and to some extent linked to the response to the preceding question, "Would like to shift more of my time to research, away from GL management".

Table 6. *Distribution of respondents by the time dedicated to GL (n=59)*

Research	Number	%
• No longer involved	20	33.9
• Involved	39	66.1
-20%	20	51.3
20% - 50%	19	48.7
Management		
• No longer involved	25	42.4
• Involved	34	57.6
-20%	16	47.1
20% - 50%	16	47.1
> 50%	2	5.8

Looking in detail at the extent of their involvement, Figure 1 illustrates how respondents split their time between research and management activities. Amongst the latter, diffusion by means of catalogues and databases of GL produced in house predominates.
There is an equal involvement in production of scientific articles and greater involvement in national rather than international conferences.

Figure 1. Distribution of respondents to the question: "Given your interest in GL , indicate time spent on the following activities"



How did respondents come to be involved in GL and their motivations

Table 7 gives data regarding the motivation underlying respondents initial involvement in GL.
The data reinforces our earlier claims regarding the nature of individuals' involvement.
As is known, GL is strongly dependent on the context in which it is produced. It comes as no surprise then that the types of motivation indicated by respondents, apart from work itself (81.4%), include interest in GL produced in house (49.2%). This connexion between Gel and the producing body is further underlined by the fact that answers to the question, "Indicate the first GL that comes to mind", nearly always referred to a document originating from individual work experience.

Table 7. *Distribution of respondents by motivation*

Motivation	% (of total number of respondents)
Work	81.4
Interest in GL of his/her Institution	49.2
Participation in national GL project	40.7
Cultural interest	39.0
Participation in GL courses and workshop	35.6
Participation in international GL proj	25.4
Other	8.9
Suggested by others	6.8
By chance	1.7

Respondents' opinions on open access and on GL definition

New communications technology, above all internet, combined with the open access movement are the significant innovations that in recent years have had greatest impact on the world of information, in particular academically and institutionally produced information. GL was affected by new types of document and format and began to operate new means of GL document transfer. These factors had to be considered in the definition. Furthermore, e-print archives and institutuional repositories are by their very nature an ideal instrument for the management and diffusion of GL. Organisational structure and content are similar to those of GL and are to some extent idicative of its evolution. (Luzi, 2005; Banks, 2004). Accordingly, we used two open questions in the survey, that sought to ascertain respondents' opinions

on a) the existence or not of a connextion between GL and open access, along with a description of any such description; b) the validity of the Luxemburg (1997) definition of GL as well as that of New York (2004). Here too, participants were invited to indicate proposed redefinitions.

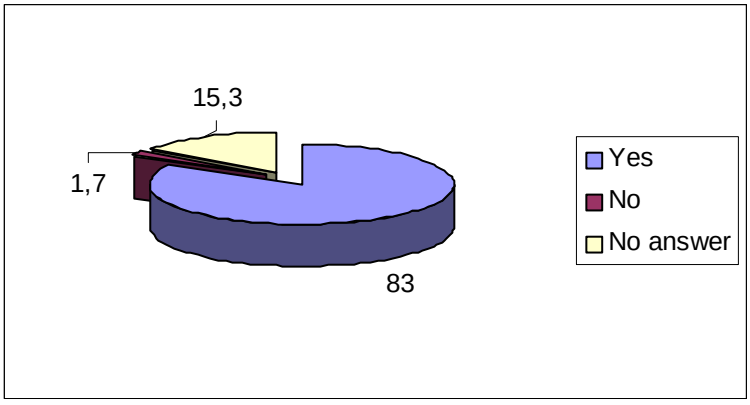
Table 7 above, *Distribution of respondents by motivation* shows that 39% of respondents claim to have come into contact with GL as a result of developing an interest in
the field. The strength of this personal motivation is highlighted by their being prepared to give their opinions on areas such as open access and the validity of the definition of GL.

a. <u>New technologies and Open access</u>

If the response regarding the contribution of IT solutions (above all internet) to removing obstacles to broader availability of GL documents came as no surprise (91.5%), the same cannot be said regarding the response (83.0%) on the connexion between Open Access and GL. It is worthy of note that a good number of respondents, (15 out of 59), clearly state what they think of this connexion. Some indicate this link to be a question of document type, “Open access is often unpublished literature or pre-print first drafts”, or in the scope “both seek to promote and spread in house scientific publications”. Others see the connexion as the means of production. “Because GL has aways been produced on a no-profit basis, and it was the first literature to appear in institutional archives with unrestricted access. To some extent it was the forerunner of Open Access. Nowadays, GL issues can be seen in the broader context of Open Access. To give one example, quality is fundamental for both types of literature, but more so for GL, which is not normally subject to mechanisms such as peer-review, even though institutions have always sought to guarantee the quality of contents. Perhaps it can be said that with Open Access, the boundary between grey and white literature is less distinct. These days GL can easily be obtained and is abundantly quoted and utilised. In the same way as open literature it an be found in on-line institutional archives”.

Taking into account the average age in the survey, along with the fact that some respondents (6) had retired, the data is indicative of their attention to the evolution of GL. The focus of this attention is primarily on new information technology, scientific communication and open access.

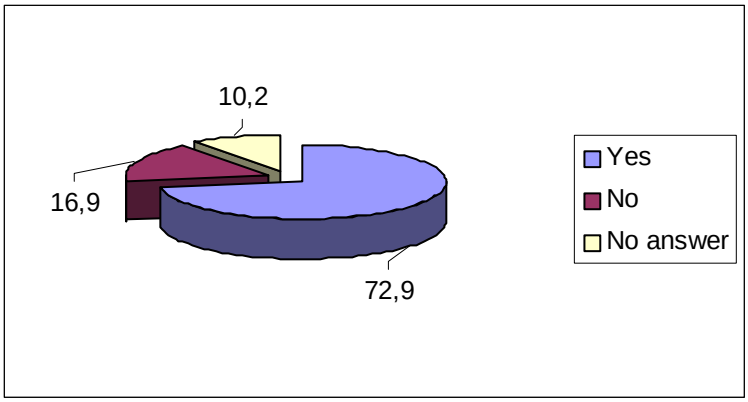
Figure 1. Distribution of respondents to the question*: Do you think that Open Access and LG are linked?*



b. <u>Definition of GL</u>

72.9% of respondents believe that the 1997 Congress of Luxemburg definition of GL, expanded on in New York in 2004, is still valid today. Other respondents, whilst still adhering to this definition, stress the need for it to include new categories of documents (“numerous musical documents, contemporary compositions, in particular of refined music”) and of document producers (“local communities”), but consider the definition has retained its validity, being grounded on the concept of information that is produced and divulgated via non commercial channels.
A practically identical result (72%) was reached in an online survey contemporary to GL6, (Boekhorst, Farace, Frantzen 2005), where participants were also asked their view on the Luxemburg definition.

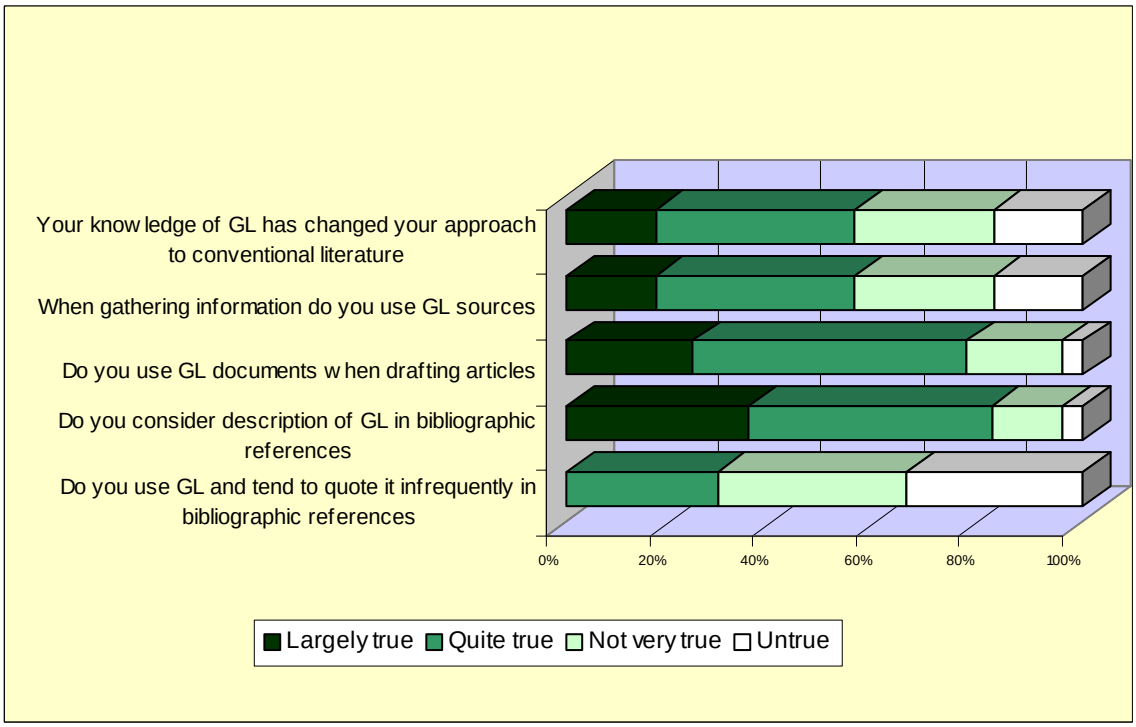
Figure 2. Distribution of respondents to the question*: Is GL definition given in Luxemburg (1997 still valuable ?*



Respondents' evaluation of their experience of GL

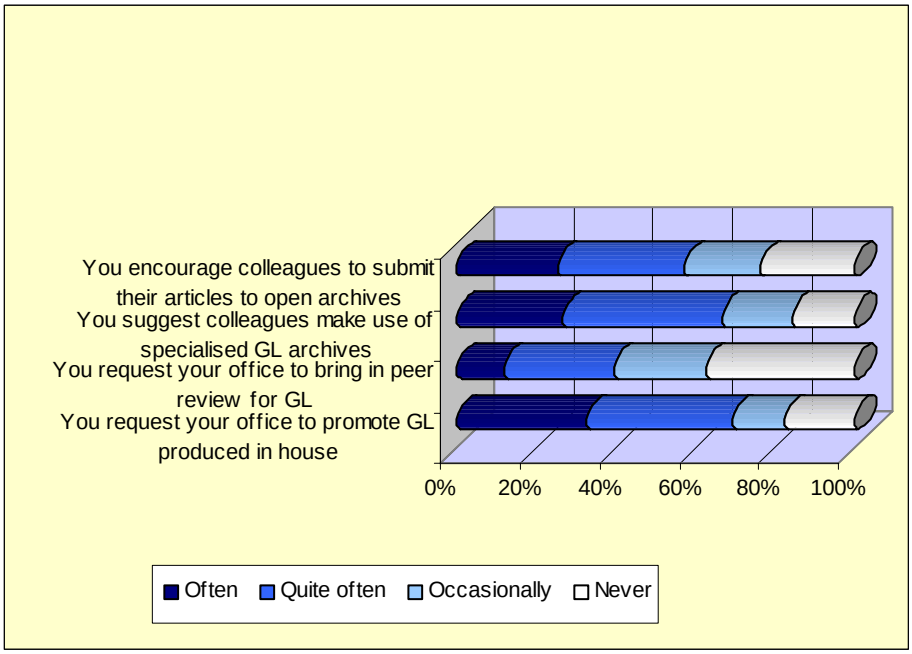
We sought to ascertain how respondents involvement with GL has modified their approach to using it, not overlooking their views on specfic problem areas.
Respondents recognise the value of GL as an information source and consequently use it while drafting their articles (See Fig. 3). Nevertheless, they are aware of the underlying problems, above all regarding the description of GL documents in bibliographic references. A correct description is a fundamental premise for the identification of a document and its correct quotation. The final question throws up some particularly interesting data. The question sought to analyse the widespread tendency to use GL and quote it infrequently.
The responses confirm respondents awareness – GL is used and quoted by 23%.

***Figure 3. Distribution of respondents to the question*: "Considering your overall experience can you determine whether….."**



Aware of the value of GL, respondents focus above all on the promotion of GL produced in house, and this confirms the above data on choices regarding GL. Respondents suggest colleagues make use of both specialised GL archives for their publications and of open archive GL banks. To a lesser extent they seek peer review for GL produced in house (See Fig. 4).

Figure 4. **Distribution of respondents to the question: Considering your overall experience can you indicate how often…..”**



In order better to ascertain respondents feeling and insight in their experience of GL, we asked them to respond to the following question, “Indicate the first GL document that comes to mind”. Instead of indicating a specific document type, the majority mentioned the most significant document in their own experience.
The question had been formulated so as to permit a variety of responses and was intended to lend itself to personal interpretations. Some of the respondents (less than half) gave a technical interpretation to the question. Their answers were technical reports, theses and reports, respectively in 18, 2 and 2 cases. All the other respondents (30) indicated a single, specific document.

The two final questions in the questionnaire sought to assess the extent to which Italian GL experts felt satisfied with their experience. Accordingly, we asked them to describe their experience with two adjectives.
This is a frequently used technique in qualitative surveys and enables the immediate recording of respondents’ impressions of the matter in hand. 79.2% replied correctly with two adjectives. 20.8% failed to respond.

As can be seen in Table 8, *interesting, stimulating, tiring useful* and *demanding* are the adjectives respondents used most frequently.

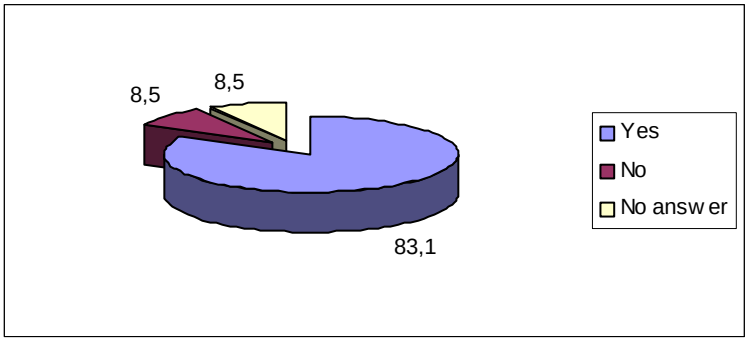
Table 8. ***Expressions used by respondents to describe their experience in GL, ranking by frequency***

Adjective	Frequency
Interesting	16
Stimulating	10
Tiring	8
Useful	7
Requiring engagement	4
Formative	3
Frustrating	3
Complex	2
Comprehensive	2
Important	2
Enriching	2

Two considerations emerge from analysis of these answers. Even though respondents state the positive aspects of their experience (stimulating, innovative, enriching etc), they do not omit to stress the hard side to working with GL, (hard, frustrating). Some openly attribute this fatigue to "inertia of my organisation (stimulating [the area] frustrating [the institutional behaviour])"; "stimulating [the novelty of the project], exasperating [at times for the inertia of my own office])".

Nevertheless, their positiveness is clear in responses to the question, "Would you choose to work in GL again?" No fewer than 83% replied affirmatively (See Fig. 6). Some respondents expanded on this answer: "Not working with GL means losing a lot of interesting material" or "More valuable than traditional printed publications". Others referred to their own knowledge, "An interesting sector to study, that stimulates across the range knowledge of various sectors of documentation", some even justifying the choice in that, "It constitutes a major challenge".

***Figure 6. Distribution of respondents to the question*: Would you choose to work in GL again?**



CONCLUSIONS

The situation emerging from the survey is on the whole positive. This goes for both respondents' involvement in GL over time, and for their enthusiasm regarding their experience.
Turning to the responses that were provided: not only were a satisfactory number of questionnaires returned (54.6%), but the quality of these responses was noteworthy. Answers were accurate and complete and respondents were prepared to explain and
expatiate on their past and present experience of GL research and management and to give their opinions on questions such as a definition of GL and the link between GL and the open access movement. The results of this qualitative survey indicate a widespread interest for the themes being surveyed. This finding is given credence by the high level of respondents' qualifications and professional position.

Many respondents are still actively involved in GL, with a prevalence working for public research institutes. In our survey, there are numerous respondents from this sector, unlike in the above-mentioned Boekhorst and others survey (Boekhorst, Farace, Frantzen 2005). Respondents to this survey were equally split between the university and public research institute sectors. The proportion in our survey can be explained in the light of the specific conditions applying in Italy. Indeed in Italy, it is these public research institutes, such as the National research council (CNR) – where the SIGLE national reference centre was located - the Institute of health (ISS), the National alternative energy board (ENEA), that are the main backers of research and were the first to set up GL management systems.

It would appear then that Italian experts are still motivated, despite the shifts in the European and domestic contexts, contexts that had originally inspired them to work in GL. The European SIGLE project has by now closed down, although it is to be reorganised on the web. As for Italy, there are a few noteworthy initiatives in the hands of various councils and associations (the LEGO database for documents and Bills from the Camera dei Deputati, the lower house of parliament; the national libraries' project on doctorate theses; AIB run divulgation, courses and seminars; etc.). Despite this, it remains the case that these initiatives need to be rationalised. This had been foreseen a few years ago by Vilma Alberani, who saw the need to create "a global, shared information system for GL".

Given the interest that respondents have in their field, it is fair to assume that their skills and experience could be harnessed to produce new open access initiatives, provided that active backup from the workplace was forthcoming. Furthermore, the majority of respondents declare agreement to putting their own GL on line and indicate participation in projects on, or being directly involved in the setting up of open institutional archives or a digital library. Here, some caution is needed as in some cases, what respondents

refer to as open archives, should more strictly be termed digital libraries rather than genuine repositories. Nevertheless, given the complexity of the field, respondents' high professional skills are beyond doubt. The significance of new information technology (Internet, Web, OA) for the production and distribution of GL is clearly understood.
Taken together, these factors lead us to believe that respondents constitute a stock of players for future initiatives that are to be hoped for.

References

Alberani Vilma, Pagamonci Armida (eds.).Monographic issue on grey literature. *Bollettino d'informazioni AIB* (1987). 27 (3-4).

Alberani Vilma, De Castro Paola (1993). A cura di. La letteratura grigia. 1° Convegno nazionale. Organizzato in collaborazione con l'Associazione Itaiana Biblioteche. Roma, 4-5 Giugno 1992. *ISTISAN Congressi, 48.*

Alberani Vilma, De Castro Paola, Daniela Minutoli (1996). La letteratura grigia. 2° Convegno nazionale. Organizzato dall'Istituto superiore di sanità in collaborazione con l'istituto centrale per il catalogo unico e l'Associazione italiana biblioteche. Roma, 20-21 maggio 1996. Atti.
ISTISAN Congressi, 48.

Alberani Vilma, De Castro Paola (a cura di) (2000). La letteratura grigia: politica e pratica. 3° Convegno nazionale. Rom,a, 25-26 Novembre 1999. *ISTISAN Congressi 67*

Banks Marcus (2004). Connections between open access publishing and access to gray literature. *J Med Libr Assoc*. 92(2): 164–166.

Boekhorst, A.K. Farace, D. J., Frantzen, J. (2005). *Grey literature survey 2004. A research project tracking develpments in the field of grey literature, in Work on grey in progress*. Sixth International conference on grey literature. New York, 6-7 December 2004. Amsterdam: TextRelease

Di Cesare Rosa, Giovanni Lazzari (2000). Towards integration of information sources on grey literature: a case study. *Journal on grey literature*, 1 (4)

GL'93. (1994). First international conference on grey literature. Conference proceedings Amsterdam, 13-15 December 1993. Amsterdam: TransAtalntic

GL'95 (1996). *Grey exploitations in the 21st. Century*. Conference proceedings. Second international conferences on grey literature. Washington D.C. 4-5 October 1995. Amsterdam: TransAtlantic

GL'97. (1998). *Perspectives on the design and transfer of scientific and technical information*. Third international conferences on grey literature. Conference proceedings. Luxemburg, 13-14 November 1997. Amsterdam: TransAtlantic

GL'99 (2000). *New frontiers in grey literature*. Second international conferences on grey literature. Conference proceedings. Washington D.C. 2-3 November 1999. Amsterdam: Greynet

GL5 (2004). *Celebrating GL conferences 1993-2003*. Fifth international conferences on grey literature. Amsterdam, December 2003. Amsterdam: TextRelease

GL6 (2005). *Work on grey in progress*. Conference proceedings. Sixth international conferences on grey literature. New York, 6-7 December 2004. Amsterdam: TextRelease

GL7 (2006). *Open Access to Grey Resources*. Sevent international conference on grey literature. Conference proceedings. Nancy, 5-6 December 2005. Amsterdam: TextRelease

Luzi, Daniela (2005). La letteratura grigia e gli open archive: un legame nascosto. *Aida informazioni* Anno 23 (1-2)

Pagamonci, Armida (1987). Il SIGLE e l'EAGLE. Un percorso di politica comunitaria e di cooperazione europea in materia d'informazione scientifica. *Bollettino d'informazioni AIB*, 27
(3-4): 348-363

Pawluch Doroty (2005). *Qualitative analysis, Sociology*. In Enciclopedia of social measurement, vol. 3. Amsterdam: Elsevier

The Messy World of Grey Literature in Cyber Security

Patricia A. Erwin
Institute for Information Infrastructure Protection. Dartmouth College, United States

Abstract
The foundation of my paper is based on four observations: 1) Research is messy; 2) Traditional collection development policies are structured documents aimed at assuring a level of quality in the collection, but also to satisfy the administrative need to justify the expense of providing resources to an academic or research community; 3) Grey literature doesn't fit the formal model of scholarly communication, therefore the quality is suspect and is not adequately addressed in most collection development policies; 4) The research process and grey literature share similar attributes. Libraries, by focusing on the formal products of research, miss the fertile, albeit more messy grey literature. I will use the I3P's focus on grey literature in cyber security as illustrative of how we need a broader definition of what constitutes the 'fruits' of research.

When we think of library collections we tend to think in terms of systems, order, and prescribed ways of tending to our collections. Our academic and research libraries are bound by the orderly world of academic departments, curriculum development, and the infrastructure of academia. Our collection development policies reflect that sense of order, clearly articulating the subject areas in which we actively collect, formats to be collected, and most importantly, the quality of the resources that will grace our shelves and gain a spot in our catalogs. What a tidy world we live in.

The truth is that research is messy. It is that intersection of the serendipity, randomness, and discovery that lends excitement to the research process. The chance merging of two seemingly unrelated concepts moves research into new areas of knowledge. These early findings and concepts do not appear in the standard scholarly communications vehicles, but rather in lab notebooks, concept papers, and technical reports, i.e. grey literature. There is a perception that grey literature is of less value than resources published through the more traditional and formal models of scholarly communications. In fact it has been noted that "scientific research is recognizable as such not because of the conditions under which it is performed but because of the way it is presented and published" (Pierce 1990, p. 55).

To better support research efforts, our collections must mirror that messiness of research. Traditionally, most collection development policies have not reflected the value of grey literature in the areas of computer science, and specifically cyber security. In support of my observations, I will report on a sampling of collection development policies from the I3P Consortium members' libraries. Our members represent academic research institutions, federal research labs, and not-for-profit research organizations. While much grey literature is collected internally, either by individual researchers or as part of an organizations institutional assets, the impetus for making this research widely available is mired in financial constraints, 'ownership' issues, and an underlying suspicion by some librarians that grey literature is not very quite as valuable and other resources that have moved through the publication process.

Overview of the I3P

The Institute for Information Infrastructure Protection (I3P) is a multi-organization consortium of academic institutions, federally funded research and development centers, and not-for-profit research organizations. The Consortium was founded in 2002 to bring experts together to identify and help mitigate threats aimed at the U.S. information infrastructure.

The I3P was founded in September 2001 to help meet a well-documented need for improved research and development (R&D) to protect the Nation's information infrastructure against catastrophic failures. The Institute's main role is to coordinate a national cyber security R&D program. Through the funding of research projects, the I3P works to identify and address critical problems in information infrastructure protection.

The information infrastructure consists of technologies and capabilities for gathering, handling, and sharing information that are accessible to, or commonly depended upon by, multiple organizations, whether within a single enterprise, a critical infrastructure sector such as banking and finance, the U.S. Government, the nation as a whole, or trans-nationally. The information infrastructure, taken as a whole, is not an engineered system. It is the result of the entrepreneurial efforts and the collective genius of the nation, working to improve efficiency and provide new opportunities for people and businesses. Security was not a major consideration at its inception, and security concerns today do not override market pressures for new uses of technology or innovation, in spite of frequent mention of hackers,

criminals, and, increasingly, terrorists and nations using or planning to use the information infrastructure as a weapon to harm the United States.[1]

Overview of the Cyber Security Digital Commons Project

A core value and long standing goal of the I3P is information sharing. An early needs assessment indicated that there was a growing need for a resource that provided easy web-based access to resources, tools, and services to support the information sharing activities among I3P Consortium members, government, and industry.

Started in 2003, the Cyber Security Digital Commons project is supported by the I3P's NIST funding. The project is focused on developing a major site for finding information in the broad area of information infrastructure protection and cyber security. Our targeted audience includes researchers, librarians, students, and the public. Project goals include: 1) Facilitate information sharing among the constituencies served by the I3P; 2) Create a variety of web-based tools and services to enhance knowledge creation under the broad topics of cyber security and information infrastructure; 3) Develop a digital library of cyber security resources.

The centerpiece of the Digital Commons is the Cyber Security Digital Library. We define the Cyber Security Digital Library as first and foremost an index to the published research in cyber security. When we say published, we extend the definition to include any intellectual thought committed to a fixed medium, such as paper or electronic media. This index is composed of meta-records that either point to a resource, or indicate how the resource might be obtained. A unique aspect of the Digital Library is the emphasis we put on acquiring and providing access to grey literature in cyber security. Because cyber security is a relatively young research area, much of the early research is found in reports, technical papers, and memos, all actively sought by project staff.

Observations About Research and Grey Literature

In the course of thinking about grey literature in connection with the traditional collection development policy, there appear to be a number of similarities between the research process and the creation of grey literature. These similarities are highlighted in the following four observations.

The process of research is messy. It is not a straight linear shot from idea to marketable product, be it a widget or peer-reviewed paper. Research is messy because learning is messy. The process of questioning, exploring, researching, and posing of ideas is generally recursive. The products of the research process, often what we think of as grey literature, is also messy. Captured bits of the research process, such as story maps, workshop outcomes, and in-house technical reports, provide a snapshot of a thinking process captured during a specific timeframe.

There is also a social networking component to the research process that can not be ignored. Researchers call, e-mail, and informally talk with other researchers about their work. A key question in looking at the value of providing access to grey literature is will it be used. Is a researcher more likely to search a database of technical reports, or call their colleague for the same information? Librarians often assume that searching a database is the best way to provide a broad overview of information available on a topic via a variety of access points. What would the world be like if instead of populating our databases based on collection development policies, we became facilitators for the social research network? We could become the party animals of the research world. Obviously, this is a tongue-in-cheek view of our role in the research process, but we do spend a significant amount of time collecting and making accessible information that, from some researchers' perspective, might be seen as easily accessible via a phone call to a colleague.

When we think of library collections we tend to think in terms of systems, order, and prescribed ways of tending to our collections. Our academic and research libraries are bound by the orderly world of academic departments, curriculum development, and the infrastructure of academia. Our collection development policies reflect that sense of order, clearly articulating the subject areas in which we actively collect, formats to be collected, and most importantly, the quality of the resources that will grace our shelves and gain a spot in our catalogs. Traditional collection development policies present a straight linear approach to providing access to the formal products of research. In fact it has been noted that "scientific research is recognizable as such not because of the conditions under which it is performed but because of the way it is presented and published" [2] The standard collection development policy is a fairly structured document

[1] (January 2003) *I3P Cyber Security Research and Development Agenda.* Retrieved October 15, 2006. Web site: https://thei3p.org.
[2] (Pierce 1990, p. 55).

aimed at assuring a level of quality in the collection, but also to satisfy the administrative need to justify the expense of providing resources to an academic or research community. This is a tidy and orderly world; we, as well as our administrations, generally are comfortable with this approach to our collection development activities.

In contrast collecting grey literature poses some risks. Although the rise of self-publishing by researchers has certainly contributed to the body of available grey literature, the quality of that literature is continues to be suspect. While grey literature "often illustrates the progression of knowledge and supports the development of new published works,"[3] its' quality continues to be suspect even among researchers. "The quality of the research that appears in grey literature is highly variable, even within grey lit produced by the same organization. The value that a respected organization would bring to a grey literature preservation project is the imprimatur of legitimacy."[4] Because of the range of quality in grey literature, and the fact that it is often freely available, as oppose to peer-reviewed commercially published resources, grey literature continues to be absent from most formal academic collection development policies. We still believe on some level that in the world of information "you get what you pay for." As we know there is tremendous cost associated with stewardship over collections, be they in paper or electronic format. Why spend our precious resources on managing information that has unproven value?

To tie this all together, we know that research is messy and often involves not only the creation of data, but also has a strong social networking component. In contrast traditional collection development policies represent the formal, secure, and auditable approach to building a collection of resources. Grey literature can be messy on many levels. It comes in a variety of formats, may or may not be of high quality, and managing grey literature via a library database may be the librarian's answer to the social network aspect of scholarly communication, effective but seen as possibly a little off the mark.

I3P Approaches to Grey Literature

Much of the unique information about cyber security falls into the category of grey literature and as such has not been commercially published or made widely available to the research community. It includes resources such as training materials, workshop notes, and technical reports. The academic papers published through such organizations as IEEE are readily available for a price. It is these other cyber security resources that the Cyber Security Digital Commons staff are most interested in making accessible.

As stated previously, the I3P is a Consortium that supports research in cyber security. The I3P has funded two multi-institutional research projects that target process control systems and the economics of cyber security. The research topics were selected by the Consortium, and were selected based on existing gaps in national research efforts, the criticality of the topic, and the impact the I3P could have through sponsoring research in the area. The research is undertaken by teams composed of multiple Consortium members, which is a hallmark of all I3P research projects. Additionally, the I3P has sponsored a number of workshops on cyber security topics. Products from these research projects and workshops included scholarly papers published in peer-reviewed journals, I3P technical reports, workshop transcripts, story maps, and fact sheets.

Researchers working on I3P-sponsored projects come from a number of different types of organizations. While the academic researchers are concerned with publishing in peer-reviewed journals, there are certainly other approaches used by the researchers to getting the fruits of their research, as well as their early thinking, out to the public. The public in this case may not mean the general public, but the audience who would be most interested in the research, including industry and government representatives. The I3P 'publishes' a series of technical reports that clearly fit into the category of grey literature. While the I3P has no formal peer review process, each of these technical reports go through a rigorous review by the project members, as well as the project researchers' home internal institutional review. In the case of the national laboratories, this review looks at both the technical quality of the research done as well as the appropriateness of the information for public release.

In looking at the collection development policies of these national laboratory libraries, the libraries are not generally collecting technical reports produced by a number of researchers working on a Consortium-funded project. They do, however, actively collect the technical reports produced in-house by their own researchers. Their approach to managing grey literature mirrors the thinking that grey literature "usually is available through specialized channels and may not enter normal channels or systems of publication,

[3] (February 2006) *Health Services Research and Health Policy Grey Literature Project: Summary Report*. Conducted by Academy Health Funded by the National Library of Medicine. Retrieved on November 1, 2006, Web site: http://www.nlm.nih.gov/nichsr/greylitreport_06.html
[4] Ibid.

distribution, bibliographic control, or acquisition by booksellers or subscription agents."[5] Searching one national laboratory library catalog for I3P funded publications yielded no bibliographic records for the documents.

On our own campus the I3P technical reports have also been ignored. While there is nothing to say we couldn't try to have these reports filtered through the library's cataloging department, they are not being actively sought out, nor would they be considered a high cataloging priority.

I3P Consortium members from not-for-profit research centers have libraries that support the research activities of the organization. Their collections databases are generally not open to anyone outside the organization. While they collect and process formal reports and technical papers these materials are considered proprietary or are produced under government contract. The government may chose to publicly release these reports, but there is no guarantee. Individual researchers may discuss their work in general ways, but specific information about projects and research is much more closely guarded.

To summarize, the I3P Consortium membership is represented by national laboratories, academic institutions, and not-for-profit research organizations. How these organizations manage grey literature represents a full spectrum of approaches, from carefully guarded nondisclosure of the information, to simply providing shelf space and little or no cataloging. The organizations are not actively collecting or cataloging I3P Consortium produced research information.

Management of Grey Literature Produced Under the Auspices of the I3P

Under the terms and conditions of the DHS award that supports I3P-sponsored research, Dartmouth College, as the recipient of the award is required to send to DHS all publications and presentations thirty day s prior to their public research. DHS does not review the information. These terms and conditions have been a produced a built-in process for collecting and managing grey literature produced by our research teams. There is an established process where research publications flow into the I3P Administrative Office and then on to the Digital Commons staff. They in turn catalog the item and make it publicly accessible, when allowable by copyright.

The Digital Commons Project staff actively collect, preserve, and make accessible (where copyright allows) these fruits of the I3P-sponsored research. The work of providing immediate accessing and long-term preservation of these resources has been left to the Digital Commons Project staff. We worry that after December 2007, when funding runs out, what will happen to these reports and corresponding bibliographic records. Over the next year we will be actively looking for either sponsorship of the Cyber Security Digital Commons, or a new home.

As part of our development of the Digital Library, the project staff is also supporting researchers working on I3P-sponsored research projects through the identification, digitization, and cataloging of pockets of historically valuable resources in cyber security that have never been commercially published or widely available to the research community.

The UC Davis Computer Security History Project

In 2003 University of California at Davis' principal I3P Consortium representative, expressed interest in having his collection of information on cyber security preserved and made accessible via the I3P Digital Library. Efforts to seek outside funding to digitize, preserve, and make accessible these resources had been unsuccessful. The researcher was convinced that the resources were valuable research documents, illustrating the early thinking and preliminary research in cyber security. In looking over the materials, and having a faith in the researcher's assessment of the value of the information, the Digital Commons staff took on the project.

Boxes of paper documents are mailed from California to New Hampshire. The documents are scanned, OCR software is run against the documents, the digital versions are reviewed for accuracy, cataloged, and then made accessible via the Digital Library. The paper documents are then returned to California. At the end of the project UC Davis will receive a CD of the documents, and electronic copies of the cataloging records. The I3P retains the right to provide access to the resources.

In many ways the actual work of the project has been very easy. It has been the knotty issues relating to grey literature that have proven to be the most challenging. Primary among them is accessing the long-term research value of the resources. We have had to take on faith that these materials are of long-term

[5] (18 January 1995) *Gray Information Functional Plan.* Report from U.S. Interagency Gray Literature Working Group. Retrieved on November 1, 2006. Web site: http://www.osti.gov/graylit/whatsnew.html

research value. Secondly, there have been copyright issues associated with the resources. Because of the age of most of these documents determining copyright ownership could be very difficult. The project prospectus made the following assumptions:

- If the item was produced as a research report (or like item) from UC Davis, it is assumed that UC Davis owns the copyright.
- If the item was produced under a contract with the Federal government and publicly releasable - We may freely convert and link to the item.
- Other items of questionable ownership – I3P staff make a good faith effort to track down copyright, but if it is not easily determined we will convert and link to the item. A publicly viewable note will be placed in the bibliographic record indicating that ownership of the item was not determined.

The UC Davis project is significant on several levels, but primarily as a lesson for the future. It was only through luck and the researcher's benign neglect of his files that the collection survived in tact. He believed the information was valuable, but could just as easily have decided it was no longer useful and thrown it away. No one in 1980 anticipated the value of this pocket of information except the researcher. His attempts to get external funding to digitize and make accessible the information were not successful. Yet, this is information that is valuable to understanding the development of field of cyber security.

This example represents a challenge for libraries. On one hand there might be grey literature resources out there that will prove to be highly valuable to future researchers. At the same time with all the competing priorities for resources we already have, how can we see into the future and know what grey literature will be valuable and what information should die a quiet death. The answer is, I believe, we can't.

New Model for Collection Development Needed for Cyber Security

The Cyber Security Digital Commons Project has taken a very different approach to systematic collection building. We acknowledge that we do not have the limits and restrictions imposed on us by an academic institution. At the same time we have a limited time span and financial resources to prove our worth. We have made a conscious decision in our collection development policy to focus on the grey literature associated with cyber security. While we continue to create and harvest meta records reflecting the more traditional products of research, such as books and journal articles, we have a special emphasis on making cyber security grey literature available.

Much of this cyber security grey literature is being produced by members of the I3P Consortium and deposited with us. We recently began a service to provide copies of our technical reports and related materials free upon request. The service has proved to be very successful. Other researchers, particularly in industry, are requesting these documents on a daily basis. We have interpreted this as a sign that these reports have immediate value.

The hope for long-term access to cyber security grey literature rests in developing research repositories and publicizing their existence. Perhaps grey literature will never attain the same status among academics that peer-reviewed literature has, but we may need to shift our world view to accepting that outside of the academic universe grey literature plays an important and legitimate role in the research process.

When we began building the Cyber Security Digital Library, there was a fair amount of skepticism about the value of some of the materials we were actively collecting and cataloging. It certainly didn't look like what faculty usually published. Our critics were right-much of what we were providing access to would not be considered on the same par as a peer-reviewed journal article. Yet, it was proving to be very popular and valuable among industry researchers. We have come to the conclusion that value is relative and that our role in the research process is to continue to actively collect, preserve, and make accessible these resources. We leave it to the researchers to ascribe value. "A researcher will look at both the grey and published literature to find the maximum amount of relevant material on a particular topic. … "Use of information demonstrates its value."[6]

Conclusion

To better support research efforts, our collections must mirror that messiness of research. Traditionally, most collection development policies have not reflected the value of grey literature in the areas of computer science, and specifically cyber security.

While much grey literature is collected internally, either by individual researchers or as part of an organizations institutional assets, the impetus for making this research widely available is mired in

[6] (1997) Grey and Black on the Internet-Interview with Dominic Farace. *CBE InternViews*, 20(3) 108-109

financial constraints, 'ownership' issues, and an underlying suspicion by some librarians that grey literature is not very quite as valuable and other resources that have moved through the publication process.

The Cyber Security Digital Commons Project, one of three projects currently supported by the I3P, has developed a collection development approach that focuses on the fruits of research, regardless of their lineage. These resources might be peer reviewed, or they may be the products closer to the research process, i.e. the grey literature. The grey literature produced through I3P-sponsored projects, including story maps, technical reports, and training materials, may not fit the traditional model, but they represent unique materials, and most importantly appear to be of most value to researchers outside the academic community.

In thinking about how to provide access to cyber security grey literature we also acknowledge the social networking aspect to research. Researchers are just as likely to call a colleague for a copy of a report than to search a library database, particularly if they assume the library is not collecting non-commercially published reports. A primary focus for our project in the coming year will be to publicize and promote the use of the Cyber Security Digital Library.

Finally, there is the question of long-term value of the materials we are so actively preserving and making accessible. We acknowledge that future value is hard to accurately predict. "The best way to predict the future is to create it."[7] We can say with some certainty that we do not if we do not provide stewardship over cyber security grey literature, there is a very good chance it will not be available to future researchers.

[7] Anonymous

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

A Public Health Knowledge Management Repository that Includes Grey Literature

Debra Revere, Paul F. Bugni, Sherrilynne Fuller
University of Washington, Seattle, WA, 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 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.

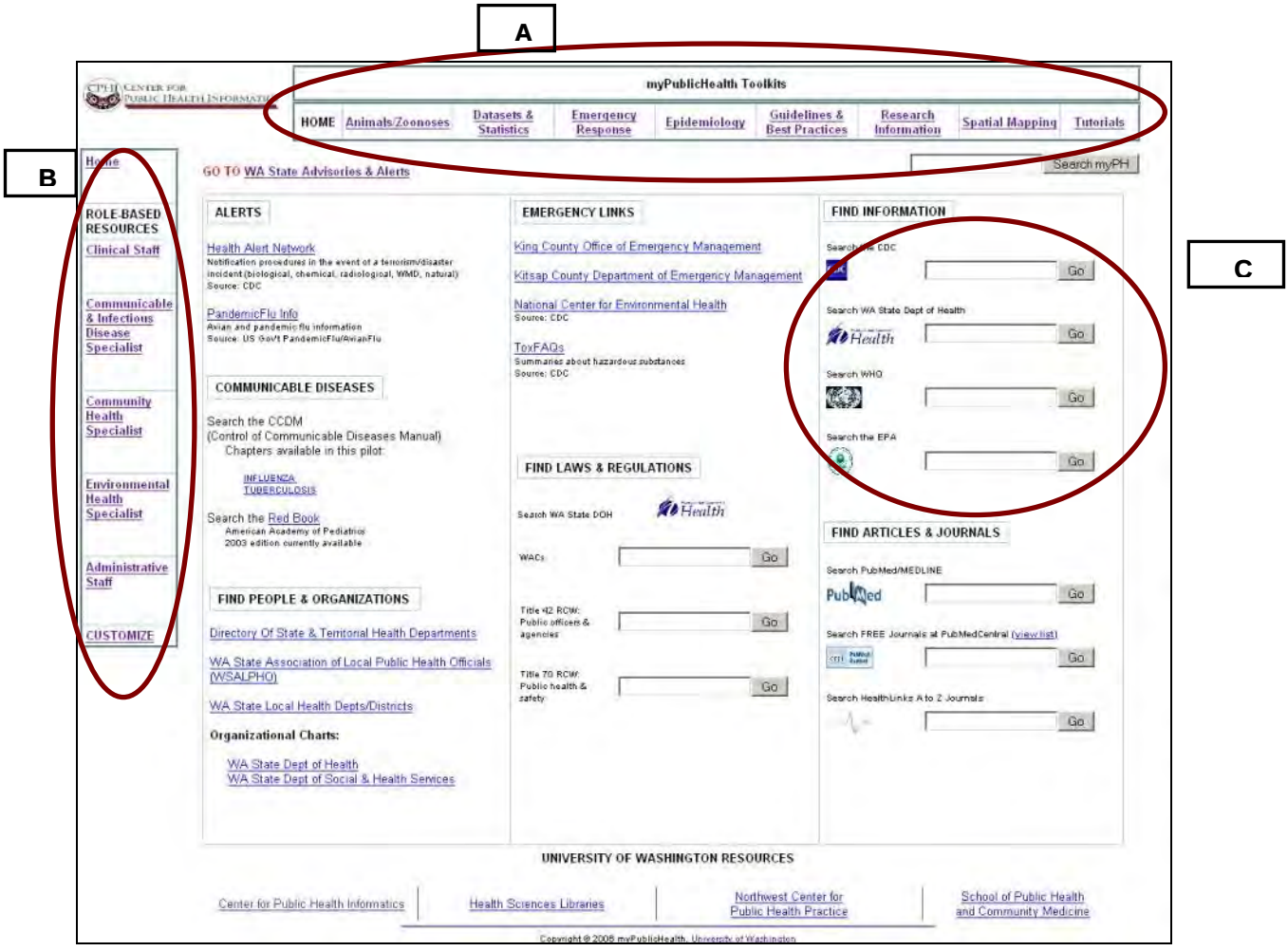


Figure 1. myPublicHealth Prototype, v.2.3

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>.

ERIC[SM] Education Resources Information Center

The Source for Grey Literature in Education

More than 100,000 open access full-text documents including:

- **Research reports sponsored by the U.S. Department of Education**
- **Conference proceedings and papers**
- **State education agency reports**
- **Research and technical reports from scholarly societies**
- **Policy papers, working papers, position papers**
- **Other education-related materials**

ERIC provides:

- **More than 1.2 million bibliographic records of documents, journal articles, and books**
- **Education grey materials indexed from the 1960s to the present**
- **A robust search interface to aid the discovery of materials**
- **Access to the online *Thesaurus of ERIC Descriptors***

Search the extensive ERIC Collection at ***www.eric.ed.gov***

ERIC is a program of the U.S. Department of Education, Institute of Education Sciences, National Library of Education.
This is a public service announcement from the U.S. Department of Education.

ies INSTITUTE OF EDUCATION SCIENCES

APPENDICES

GreyNet Grey Literature Network Service

Corporate Authors and Associates

ASSOCIATE MEMBERS ARE THE CORPORATE AUTHORS OF GREYNET'S PRINT AND E-PUBLICATIONS

GreyNet's main activity lies in "Corporate Authorship" of electronic and print publications as well as web resources. GreyNet is conscious of the fact that these valuable resources are made possible through the intellectual and financial contributions of leading organisations in the field of grey literature. These organisations fulfil a role that is beyond sponsorship, because they are *pari passu* responsible for the further development of GreyNet's content base, which currently includes:

GreyNet Grey Literature Network Service
Website
www.greynet.org

GreySource
Index
www.greynet.org/greysourceindex.html

GreyText
Archive
www.greynet.org/greytextarchive.html

Whois in **GL**
Biographical Notes
www.textrelease.com/textreleasehome.whois2006

GL 8 Harnessing the Power of Grey
Conference Memoranda
ISSN 1574-1214

GL 8 Harnessing the Power of Grey
Conference Program and Abstracts
ISSN 1385-2308

GL 8 Harnessing the Power of Grey
Conference Proceedings
ISSN 1386-2316

the grey Journal
TGJ Volume 2, 2006
ISSN 1574-1796

Recognition of the support of these Associate Members is duly manifest here and on all GreyNet's print and electronic publications.

GreyNet Associate Members 2006

NLE, United States
National Library of Education;
U.S. Department of Education

INIST-CNRS, France
Institute for Scientific and Technical Information;
National Centre for Scientific Research

NYAM, United States
The New York Academy of Medicine;
Division of Information Management

List of Participating Organizations

American Psychological Association, APA/PsycEXTRA	United States
Arizona State University Libraries, AZU	United States
Battelle/CBIAC	United States
British Library, BL	United Kingdom
Centre National de Recherche Scientifique, CNRS	France
CCLRC - Rutherford Appleton Laboratory	United Kingdom
Dalhousie University, DAL	Canada
Dartmouth College; I3P	United States
East View Information Services	United States
EBSCO Information Services	United States
Federal Library and Information Center Committee, FLICC	United States
Federal Library and Information Network, FEDLINK	United States
Fisheries Library; Food and Agriculture Organization, FAO	Italy
Flemish Department of Economy, Science and Innovation	Belgium
Fred Hollows Foundation	New Zealand
Global Library Services Network, GLSN	Australia
Grey Literature Network Service, GreyNet	Netherlands
Information International Associates, IIa	United States
Institut de l'Information Scientifique et Technique, INIST	France
Istituto di Ricerche Popolazione e le Politiche Sociali, IRPPS	Italy
Istituto Superiore di Sanità, ISS	Italy
Japan Science and Technology Agency, JST	Japan
Labriola American Indian Data Center; ASU	United States
Library of Congress, LC	United States
Library of Zeeland, ZEBI	Netherlands
Louisiana State University, LSU	United States
NASA-GFSC, Goddard Space Flight Center	United States
National Library of Education, NLE	United States
National Library of Medicine, NLM	United States
National Research Center, CNR	Italy
New York Academy of Medicine, NYAM	United States
New York University School of Medicine, NYU	United States
Office of Scientific and Technical Information, OSTI	United States
Oregon State University	United States
Pacific Open Learning Health Net; WHO	Fiji
Portland State University, PSU	United States
United States Department of Energy, DOE	United States
University of Bergen, UiB	Norway
University of Bucharest, UNIBUC	Romania
University of California, Irvine Libraries, UCI	United States
University of New Orleans, UNO	United States
University of South Florida, USF	United States
University of Washington; CPHI	United States

Author Information

Nikkia Anderson is an 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. At the NASA Goddard Space Flight Center Library, she manages some digital collections and coordinates on various knowledge management initiatives, which involve bringing some grey areas to light. 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. For several years she was the chairwoman of the Steering Group of the national CRIS system and project secretary of a National system for academic administration. Anne is presently representing UiB in the national group that is implementing a new national research documentation system, FRIDA. She also participated in The CORDIS funded European-wide project on " Best Practice" 1996 and was a member of the working group set up 1997 that produced the report CERIF2000 Guidelines (1999) www.cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS www.eurocris.org is now the custodian of the CERIF model and Anne is a member of the euroCRIS Board. She has published in the areas of CRIS and Grey Literature and served on programme committees for the GL conference series.
Email: anne.asserson@fa.uib.no

Marcus A. Banks is a librarian at the New York University School of Medicine. He is also the editor of the open access journal Biomedical Digital Libraries, and the Chair of the Medical Library Association's Task Force on Librarians without Borders. He is also the recipient of the GreyNet Award 2006. He is interested in how the concept of grey literature will continue to evolve in the digital age.
Email: banksm01@library.med.nyu.edu

Cees de Blaaij studied Social and Economic History at the University of Nijmegen and Library Science at the University of Amsterdam. He worked for Ernst & Young, management consultants, and the Institute for Information Law (University Amsterdam) as information professional. At the moment he is working as an academic librarian and coordinator digital services for the Public and Academic Library of Zealand, Netherlands. He took part in several GL conferences. He published several articles on issues concerning copyright in the digital environment and accessibility of grey literature on the Internet. Email: cdbl@zebi.nl

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

Jean Collins is Fisheries Librarian at FAO where she has worked since 1990 providing information services primarily to the FAO Fisheries Department and fisheries organizations in developing countries. She is very active in the International Association of Aquatic Libraries and Information Centers (IAMSLIC) strengthening professional network among fisheries libraries. She is also editor of several FAO publications addressing the issues of libraries in developing countries and the role of information in fisheries development. Email: jean.collins@fao.org

Courtney Crummett is a fellow at the National Library of Medicine in Bethesda, Maryland. She holds an MA from University of South Florida School of Library and Information Science and an MS in geology from the University of Maryland. Her research interests include usage of electronic tools such as, blogs, RSS feeds, wikis, to facilitate information seeking for scientists.
Email: courtcrum@gmail.com

Paola De Castro is responsible for the production and diffusion of open and grey literature issued by the Istituto Superiore di Sanità, the Italian National Institute of Health, producing a quarterly official journal, a non-commercial monthly newsletter and different series of technical reports. She has published many articles on the information transfer process both at national and international level with special reference to the role of grey literature and takes part in the national and international debate in the field; she delivers courses on scientific writing for the Italian National Health Service operators, she is an active member of the European Association of Science Editors (EASE), the European Association of Health Information and Libraries (EAHIL) and the Italian Library Association (AIB).
Email: paola.decastro@iss.it

Rosa Di Cesare was born in Civita d'Antino (AQ) in 1952 and graduated from "La Sapienza" University in Rome in 1982. She received her diploma in Librarianship from the Vatican Library in 1996. She worked in the Central Library of National research council (CNR), where she started to become involved in research activity in the field of Grey literature (GL). Member of the Technical Committee for the SIGLE database from 1995 to 2001, she is presently responsible for the Library at the Institute of research on population and social policies (IRPPS) of the National research council. Her studies have focused on citation analysis and on the use of GL in scientific publications. Email: r.dicesare@irpps.cnr.it

Patricia Erwin is Senior Assistant Director for Informatics Services at the Institute for Information Infrastructure Protection (Dartmouth College, Hanover, New Hampshire). She has held information management related positions at the Dartmouth College Library, the University of Virginia Library System, and in private industry. She regularly teaches graduate and undergraduate courses in digital libraries, knowledge management, organizational learning, and critical thinking. Ms Erwin holds a Masters Degree in Library and Information Science from Indiana University, and a Masters Degree in Internet Strategies Management from Marlboro College. Her primary interests are in how organizations leverage what they know to push a disciple or area of knowledge forward, the role management of information plays in the research process, and how organizational structure enhances or impedes the flow of information. Most recently she has been directing a project to develop a national digital library commons of cyber security resources. This project is funded by the U.S. Department of Commerce's National Institute of Science and Technology (NIST). Email: perwin@thei3p.org

Author Information

continued

Dominic J. Farace is Director of TextRelease, an Amsterdam based information bureau specializing in grey literature and networked information. He is a native Louisianan and holds two degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked for the past twenty-seven years. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (Swidoc/KNAW), he founded GreyNet, Grey Literature Network Service, in 1993 and has since been responsible for the international GL-Conference Series. In this capacity, he serves as Program and Conference Director as well as managing editor of the conference proceedings. Since 2004, he is a Guest Lecturer on Grey Literature in the Masters Program at the University of Amsterdam; Guest Editor for PRQ, Publishing Research Quarterly and Editor of TGJ, The Grey Journal. Email: info@greynet.org

Jerry Frantzen graduated in 1999 from the College of Amsterdam in Library and Information Science. He is a freelance information specialist and the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet, Grey Literature Network Service. Email: jerry@frantzen.eu

Julia Gelfand has been a librarian with the University of California, Irvine Libraries since 1981. She has been tracking the grey literature movement since the late 1980s and has participated in all of the previous GL conferences and has published and presented widely on different topics in grey literature. Her particular interests are in scholarly communications, electronic publishing, collection development, bibliography of science and technology, and she thinks that with more emphasis on networking and digital libraries, Grey Literature has a very interesting future. She is currently the chair of the IFLA Science & Technology Section and vice-chair/chair-elect of the ALA ACRL Science & Technology Section. Email: jgelfand@uci.edu

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

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 CCLRC (Council for the Central Laboratory of the Research Councils), based at Rutherford Appleton Laboratory in UK. This involves coordinating IT development across multiple research fields both within the organisation and outside and also coordinating the international relations of the organisation with other similar organisations. Previously he led the 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. Keith is a Fellow of both the Geological Society of London and the British Computer Society. He is a Chartered Engineer and Chartered IT Professional. 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. Keith 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 IT support of relationships between hypotheses, experiments, primary data and publications based on research in information systems, knowledge-based systems and metadata. Email: k.g.jeffery@rl.ac.uk

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

Bertrum H. MacDonald is the Associate Dean (Research) in the Faculty of Management at Dalhousie University, Halifax, Canada. He holds a BSc (Biology) from Acadia University, and an MA (History of Science), MLS, and PhD (Information Science) from the University of Western Ontario. He is the Editor of Electronic Resources for the History of the Book in Canada / Histoire du livre et de l'imprimé au Canada project www.hbic.library.utoronto.ca>, funded by a $2.3 million grant from the Social Sciences and Humanities Research Council of Canada. His research focuses primarily on the dissemination of scientific information among Canadian scientists and engineers (both historically and currently) and on the history of scientific and technical publishing in Canada. In 2001 he was named a Dibner Library Research Scholar at the Smithsonian Institution in Washington DC. In recognition of his significant contributions to the bibliography of Canadian science and technology he was awarded the Marie Tremaine Medal by the Bibliographical Society of Canada in 2000 and is the GreyNet Award Recipient in 2004.
Email: bertrum.macdonald@dal.ca

Author Information

continued

Patricia Merrikin is Reference Librarian at the David Lubin Memorial Library of the Food and Agriculture Organization (FAO) of the U.N. where she has worked since 1992. Her subject areas include Agriculture, Rural Development as well as Development economics and sociology. She is one of the key contacts for AGLINET, an international network of agriculture libraries. Ms Merrikin works closely with FAO staff providing expert assistance with information identification and retrieval.
Email: patricia.merrikin@fao.org

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.
Email: 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

Sandra Salinetti deals with the production of technical reports produced by the Istituto Superiore di Sanità, assisting authors in the preparation and editing of documents and reviewing the texts to comply with national and international standards and make them available through the Internet. She participates in a training project on scientific writing to improve authors' personal skills in the production of scientific documents with special reference to the presentation of technical reports. Email: Sandra.Salinetti@iss.it

Joachim Schöpfel graduated from the University of Hamburg in 1984. A research assistant and lecturer at the University of Hamburg, Department of Developmental and Educational Psychology, from 1985 to 1990, he obtained his Ph.D. from the same university in 1992. He is presently head of the library department at the French Institute of Scientific and Technical Information and teaches Culture and Society (1992-2001) and Documentation (from 2001 on) at the University of Nancy. He is member of the UK Serials Group and was the last president of EAGLE.
Email: schopfel@inist.fr

Christiane Stock graduated from the University of Freiburg in 1984. She joined INIST-CNRS, the French Institute of Scientific and Technical Information in 1989. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). Today she is the head of the monographs and grey literature section at INIST. Email: christiane.stock@inist.fr

Janet Webster is the head librarian at Oregon State University's Guin Library located at the Hatfield Marine Science Center in Newport, Oregon U.S.A. 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

Peter G. Wells is a Senior Research Scientist and Research Manager with the Environmental Conservation Branch, Environment Canada, Dartmouth, Nova Scotia, Canada, specializing in coastal and marine water quality problems and issues, and marine ecotoxicology. He is a Professor and Adjunct Professor at Dalhousie University, Halifax. He is a past Chair (1998-2000) of GESAMP, which publishes a technical Reports and Studies series on marine pollution and protection issues, through the UN agencies. He has been involved in a number of state of the oceans reports, including most recently the UNEP-GESAMP Report and Studies No. 70, A Sea of Troubles (2001). One of his research interests is evaluating the use and influence of such technical reports in marine environmental management. Email: peter.wells@ec.gc.ca

Index to Authors

GreyNet Members
20% Discount

GL-Conference Series, ISSN 1386-2316

Publication Order Form

Compiled and Edited by GreyNet

	No. of Copies	*x*	*Amount in Euros*	**Subtotal**
GL8 Conference Proceedings, 2007 – Printed Edition *ISBN 90-77484-08-6 Postage and Handling excluded*[*)]		**x**	**75.00 =**	**€** (INCL. VAT/BTW)
GL7 Conference Proceedings, 2006 – Printed Edition *ISBN 90-77484-06-X Postage and Handling excluded*[*)]		**x**	**70.00 =**	**€** (INCL. VAT/BTW)
GL6 Conference Proceedings, 2005 – Printed Edition *ISBN 90-77484-04-3 Postage and Handling excluded*[*)]		**x**	**70.00 =**	**€** (INCL. VAT/BTW)
GL5 Conference Proceedings, 2004 – Printed Edition *ISBN 90-77484-01-9 Postage and Handling excluded*[*)]		**x**	**60.00 =**	**€** (INCL. VAT/BTW)

Postage and Handling per printed copy[*)]

Holland		**x**	**5.00**	**€**
Europe		**x**	**10.00**	**€**
Other		**x**	**20.00**	**€**
			Total Amount:	**€** **(INCL. VAT/BTW)**

Customer information

Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
Fax/Email Address:	

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

❑ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to "Publication Order"

❑ Bank check/draft made payable to TextRelease, with reference to "Publication Order"

❑ MasterCard/Eurocard ❑ Visa card ❑ American Express

Card No. _____ Expiration Date: _____

Print the name that appears on the credit card, here _____

Signature: _____ **CVC II code:** _____ *(Last 3 digits on signature side of card)*

Place: _____ Date: _____

Note On receipt of payment, an invoice marked paid will be sent to your postal address. If you require a *pro forma* invoice, please notify:

TextRelease
www.textrelease.com

GL Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: **info@textrelease.com**