https://doi.org/10.26069/greynet-2026-000.606-tgj
ISSN 1574-1796

An International Journal on Grey Literature

Spring 2007 - TGJ Volume 3, Number 1

'Grey Standards in Transition and Use'

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

https://doi.org/10.26069/greynet-2026-000.606-tgj

The Grey Journal
An International Journal on Grey Literature

COLOPHON

CIP
The Grey Journal (TGJ): An international journal on grey literature / D.J. Farace (Journal Editor); J. Frantzen (Technical Editor) ; GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 3, Number 1, spring 2007. - ***INIST-CNRS, NLE,*** *and* ***NYAM*** *are corporate authors and associate members of GreyNet. This serial publication appears three times a year - in spring, summer, and autumn. Each issue is thematic and deals with one or more related topics in the field of grey literature. The Grey Journal appears both in print and electronic formats.*
ISSN 1574-1796 (Print)
ISSN 1574-180X (E-Print/PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:
€100 individual, including postage & handling
€215 institutional, including postage & handling

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

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

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

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

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

About *GreyNet*, 1992-2007
The Grey Literature Network Service celebrates this year its 15th Anniversary. GreyNet was established in order to facilitate dialog, research, and communication between persons and organizations in the field of grey literature. GreyNet further seeks to identify and distribute information on and about grey literature in networked environments. Its main activities include the International Conference Series on Grey Literature, the creation and maintenance of web-based resources, a moderated Listserv, and The Grey Journal. Currently, GreyNet is engaged in the development of a distance learning course for graduate and post-graduate students.

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

https://doi.org/10.26069/greynet-2026-000.606-tgj

Contents
"GREY STANDARDS IN TRANSITION AND USE"

Grey Literature - Taxonomies and Structures for Collection Development 7
Julia Gelfand *(United States)*

Metadata-based analysis to improve clinical trial exchange 17
Daniela Luzi, Fabrizio L. Ricci *(Italy)*, and Luca Dan Serbanati *(Romania)*

Implications of Copyright Evolution for the Future of Scholarly Communication and Grey Literature 31
Marcus A. Banks *(United States)* and Cees de Blaaij *(Netherlands)*

Legal Foundations of the Scientific and Technical Grey Literature Development in Russia 37
Leonid P. Pavlov *(Russia)*

From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource Migration in European Context
Joachim Schöpfel, Christiane Stock, and Nathalie Henrot *(France)* 45

Scientists produce and use grey literature, but are they aware of the implications of doing so?
Paola De Castro *(Italy)* 52

Colophon 2
Editor's Note 5

Advertisements
- EBSCO Information Services 4
- ERIC-NLE, National Library of Education 6
- GESIS, German Social Science Infrastructure Services 36
- INIST-CNRS, Institute for Scientific and Technical Information 44

GL9 Announcement and Call-for-Papers 56
About the Authors 58
Notes for Contributors 59

https://doi.org/10.26069/greynet-2026-000.606-tgj

https://doi.org/10.26069/greynet-2026-000.606-tgj

"GreyNet, 1992-2007"

This year marks GreyNet's (Grey Literature Network Service) 15th Anniversary. Looking back in the field of information and communication to 1992 is not only a journey in time to a previous century but also to a period in history without mobile telephone, the World Wide Web, Open Access Repositories, and other such technological advances and exploitations. When we look back to this same period related to grey literature, we see an *ad hoc* interest in research among a few information professionals in Europe and the United States; however, change was imminent.

In 1993, a year following GreyNet's foundation, the First International Conference on Grey Literature was to bring together a core group of authors and researchers sowing the seeds for a global research community. The endeavor met with success in that the Ninth International Conference in the GL-Series convenes this December in Antwerp, Belgium.

Over the past 15 years, GreyNet has been able to mobilize over 250 authors and researchers providing them with a variety of platforms for presenting and communicating the results of their work. In 2005, GreyNet undertook editing of The Grey Journal (TGJ) in an effort to provide another channel for the distribution of research results in the field of grey literature; and, the journal is currently in its third volume. As of recent, GreyNet has initiated a proposal for a distance learning course for (post)graduate students. GreyNet hopes that such an innovative endeavor will not only further the awareness and distribution of research results in the field of grey literature but will at the same time provide new avenues for young researchers to enter and specialize in this field. Since The Grey Journal is geared to colleges and schools of Library and Information Studies (LIS), a distance learning course would be consistent with GreyNet's other networked information services – first conference based, later journal bound, and soon hopefully course certified.

This issue of TGJ is titled "Grey Standards in Transition and Use". In a way it is a tribute to some of the core authors and researchers in the field of grey literature, who have set standards and who worked throughout a period of rapid transition in the field of information to maintain grey literature's identity and importance. Three of the authors in this current issue are GreyNet Award Recipients: Julia Gelfand (1999), Daniela Luzi (2000), and most recently Marcus A. Banks (2006).

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

https://doi.org/10.26069/greynet-2026-000.606-tgj

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

https://doi.org/10.26069/greynet-2026-000.606-tgj

Grey Literature: Taxonomies and Structures for Collection Development *

Julia Gelfand *(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

* First published in the GL8 Conference Proceedings, February 2007

- 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) [1]

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.[2] 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[3] 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.[4] 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. [5]

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."[6] 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[7]

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[8]

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[9]

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[10]

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.[11]

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 (http://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."[12] The Visual Information Seeking Mantra of "Overview first, zoom and filter, then details-on-demand"[13] 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."[14]

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[15] 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."[16] 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[17]

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[18]

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."[19]

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."[20] 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[21]

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

[1] Cataloging Policy and Support Office, Library of Congress, *Library of Congress Classification Outline*, http://www.loc.gov/catdir/cpso/lcco/lcco.html
[2] 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.
[3] 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.
[4] ARL Statistics for 1999-2000, reported in *ARL Bi-monthly Report # 219*, December 2001 - see http://www.arl.org/newsltr/219/eresources.htm
[5] 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.
[6] *Oxford English Dictionary Online* - http://dictionary.oed.com
[7] Alan Gilchrist, "Thesauri, Taxonomies and Ontologies: An Etymological Note." *Journal of Documentation* 59 #1, 2003: 7-18.
[8] *Ibid*.
[9] *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.
[10] 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
[11] Edward R. Tufte, *Envisioning Information*. Cheshire, CT: Graphics Press, 1990.
[12] 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.
[13] *Ibid*: 337.
[14] *Ibid*: 341.
[15] 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.
[16] 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.
[17] *Ibid*: 289.
[18] Delphi Group, *Information Intelligence: Content Classification and the Enterprise Taxonomy Practice.* Boston, MA: Delphi Group, 2004:11.
[19] Bonnie Nardi and Vicki L. O'Day, *Information Ecologies: Using Technology With Heart.* Cambridge, MA: MIT Press, 1999: 49.
[20] Thomas Davenport with Laurence Prusak, *Information Ecology: Mastering the Information and Knowledge Environment.* NY: Oxford University Press, 1997: 41.
[21] *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.

https://doi.org/10.26069/greynet-2026-000.606-tgj

Metadata-based analysis to improve clinical trial exchange *

Daniela Luzi and Fabrizio L. Ricci *(Italy)*
Luca Dan Serbanati *(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 organizations, where grey literature documents have a central role in all the different phases of the complex CT process.

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.

* First published in the GL8 Conference Proceedings, February 2007

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

[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.

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.

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:

A. International regulatory bodies,
B. Organisations promoting standards, and
C. 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.

https://doi.org/10.26069/greynet-2026-000.606-tgj

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

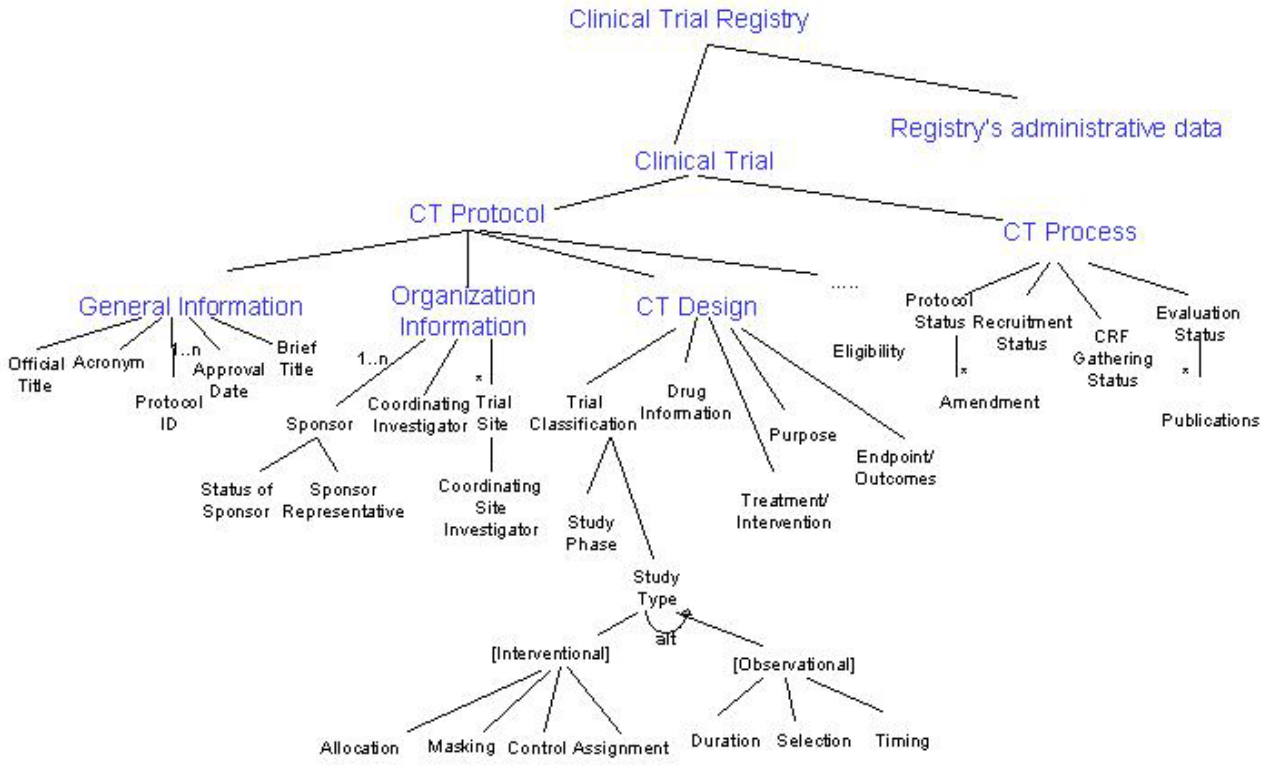


Fig. 1.- Information Structuring schema of CT registriies

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 the 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.

https://doi.org/10.26069/greynet-2026-000.606-tgj

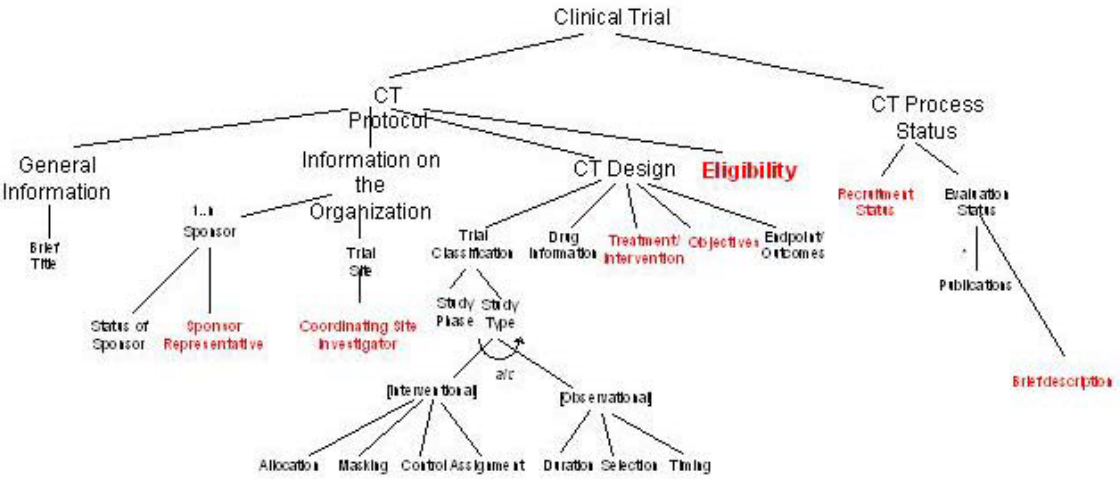


Fig. 2. - Patient's information view

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.

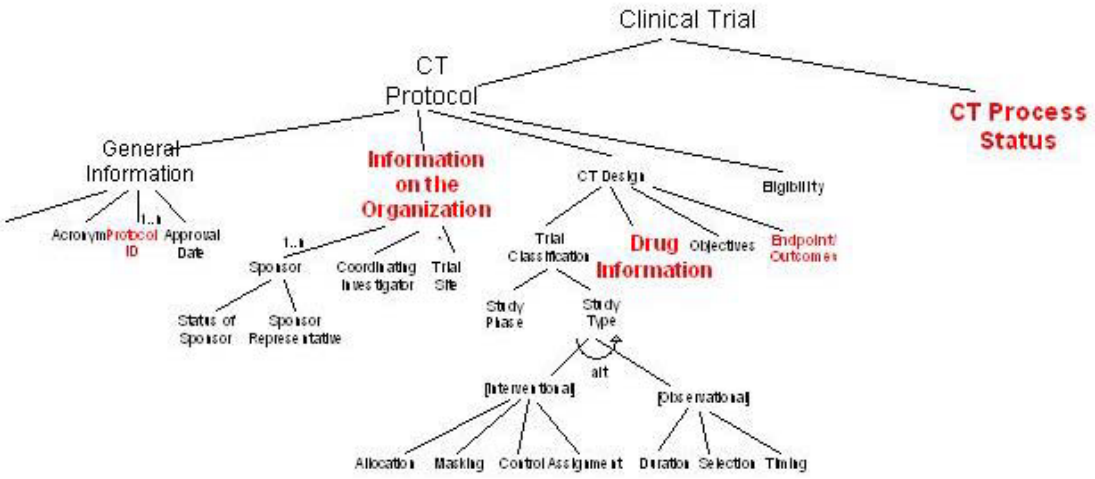


Fig. 3. - National Health Authority's Information view

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.

https://doi.org/10.26069/greynet-2026-000.606-tgj

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

Marcus A. Banks *(United States)*
Cees de Blaaij *(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.

* First published in the GL8 Conference Proceedings, February 2007

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

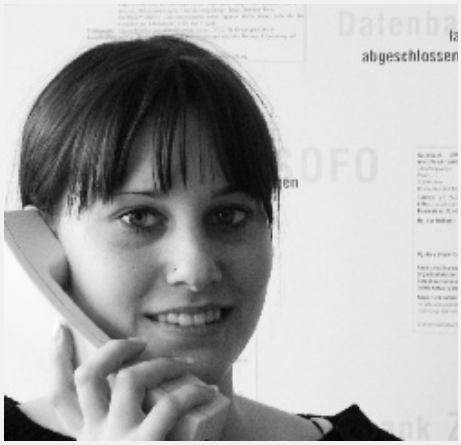
1. SHERPA/RoMEO-Publisher copyright policies & self-archiving. http://www.sherpa.ac.uk/romeo.php [Accessed 11 November 2006].
2. arXiv.og e-Print archive. http://arxiv.org/ [Accessed 11 November 2006].
3. Bethesda Statement on Open Access Publishing. http://www.earlham.edu/%7Epeters/fos/bethesda.htm [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.

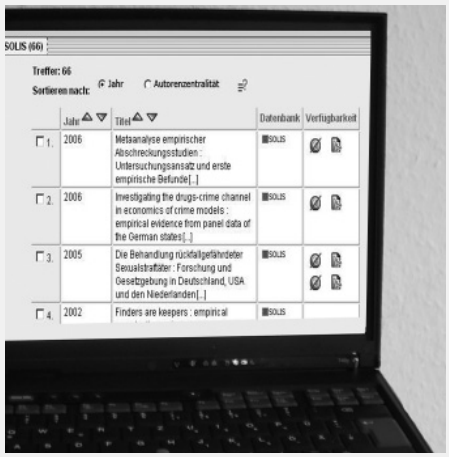
10. Senate of the United States (2006), Federal Research Public Access Act of 2006. Online at: http://cornyn.senate.gov/doc_archive/05-02-2006_COE06461_xml.pdf [Accessed 11 November 2006].
11. ATA-Federal Research Public Access Act of 2006. http://taxpayeraccess.org/frpaa/index.html [Accessed 11 November 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].

https://doi.org/10.26069/greynet-2026-000.606-tgj

gesis
Social Science Information Centre

Services on Grey Literature





... in Eastern Europe

The **gesis Service Agency Eastern Europe** supports East-West cooperation in social sciences and contributes to the networking activities in the European research area. Among other services, the agency offers consultancy on access to grey literature from Eastern Europe.

Access via: www.gesis.org/eastern_europe

... from Germany, Austria, Switzerland

Our database **SOLIS** contains more than 300.000 citations - including about 50.000 citations with abstracts to grey literature in the social sciences in German speaking countries.

Access via:
» www.infoconnex.de
» www.stn-international.de
» www.gbi.de

gesis also provides search services.

More information at: **www.gesis.org**

https://doi.org/10.26069/greynet-2026-000.606-tgj

Legal Foundations of the Scientific and Technical Grey Literature Development in Russia

Leonid P. Pavlov *(Russia)*

Introduction

In 1966 the leading Russian institution for scientific grey literature collecting, processing, storing and disseminating – the Scientific and Technical Information Centre of Russia (VNTIC) – was founded. In the sixties, it was internationally acknowledged that the Soviet achievements in nuclear power engineering and space exploration were largely supported by a well-defined scientific and technical information system. The fact that in respect of grey literature the system has survived the devastating nineties only confirms its efficiency and validity. In July 1997 the Decision of the Government of the Russian Federation N 950 (it is still well known in Russia under this number as "950th Decision") "The Regulations on the State System for Scientific and Technical Information of the Russian Federation" was taken that outlined the importance of developing an STI system on the state level and confirmed the leading role of VNTIC in the area of Russian grey literature.

In the context of recent international efforts, say, the initiative of "Nancy style" recommendations, to elaborate and promote some standards for improving the quality of scientific research results representation it seems useful to give a brief review of the state of the art in Russia, all the more that because of the language barrier the Russian documents are less known in the West. The relevant Russian official documents of different levels (federal laws, government decisions, standards, VNTIC instructions, etc.) concerning scientific grey literature, reports and dissertations representation, both full- text and metadata, are reviewed in the paper. The review is naturally divided into two parts concerning reports and dissertations correspondingly but before getting down to the subject itself a few remarks on terminology should be made since the Russian situation is being rendered in the English language.

In 1966 the leading Russian institution for scientific grey literature collecting, processing, storing and disseminating – the Scientific and Technical Information Centre of Russia (VNTIC) – was founded. In the sixties, it was internationally acknowledged that the Soviet achievements in nuclear power engineering and space exploration were largely supported by a well-defined scientific and technical information system. The fact that in respect of grey literature the system has survived the devastating nineties only confirms its efficiency and validity. In July 1997 the Decision of the Government of the Russian Federation N 950 (it is still well known in Russia under this number as "950th Decision") "The Regulations on the State System for Scientific and Technical Information of the Russian Federation" was taken that outlined the importance of developing an STI system on the state level and confirmed the leading role of VNTIC in the area of Russian grey literature.

In the context of recent international efforts, say, the initiative of "Nancy style" recommendations, to elaborate and promote some standards for improving the quality of scientific research results representation it seems useful to give a brief review of the state of the art in Russia, all the more that because of the language barrier the Russian documents are less known in the West. The relevant Russian official documents of different levels (federal laws, government decisions, standards, VNTIC instructions, etc.) concerning scientific grey literature, reports and dissertations representation, both full- text and metadata, are reviewed in the paper. The review is naturally divided into two parts concerning reports and dissertations correspondingly but before getting down to the subject itself a few remarks on terminology should be made since the Russian situation is being rendered in the English language.

Terminological digression

When describing phenomena of one realm using the language of another we must be sure to call a spade a spade. This is especially important when doing Russian-Western or Western-Russian translation, much more important than for, say, French-English or German-Dutch translation. A living language reflects nation's mentality and reality and it is well known but sometimes is underestimated in the West that the Russian mentality and reality (or vice versa, depending whether you adhere to idealistic or materialistic philosophy) are rather different from those in the West.

In 1997 when Russia joined EAGLE and VNTIC began to send documents translated into English to SIGLE database there was a problem of converting the subject heading codes assigned to the Russian documents in accordance with the State Subject Heading Classifier for Scientific and Technical Information (abbreviated in Russian as GRNTI) to the subject heading codes of COSATI-SIGLE used in the SIGLE database. I had to translate the COSATI-SIGLE classifier from English into Russian so that to establish a correspondence between the two systems of classification and coding. The difference between the classifiers was striking. It was evident that the problem of scientific subjects' taxonomy was not monosemantic and simple and the codes conversion couldn't be done formally, without a certain preliminary semantic harmonization of the subjects' names. Like the living natural language itself, the classification reflects national traditions, mentality and system of values. For example, it was quite difficult to understand why there was a separate first-level subject heading called "Ordnance" in COSATI-SIGLE, of the same level and importance as, say, "Mathematics", "Physics", etc. but at the same time some evident (for a Russian) scientific subjects (like informatics, cybernetics) were absent from the classifier. Nevertheless, using non-formal methods the correspondence was established and all the VNTIC documents prepared for SIGLE were assigned the COSATI-SIGLE subject headings.

Therefore, there is always a problem, especially when doing translation from the Russian language of juridical, economic and official texts, either to give an English term well understood in the West but partly to loose the original Russian meaning or to stick to the Russian cliché with a full spectre of the Russian specificity with the risk of not being understood by the Western readers. For example, one of the most effective directive documents in the Russian bureaucratic machine is the President's edict (or decree) should one use the words of the Latin origin, quite clear for any Western reader. However, in Russian it is called President's "ukaz", included in the English language dictionaries and spelled in English like "ukase" which is richer and more precise in meaning but needs a certain interpretation for a common English-language reader.

In the next sections of the paper I will try to bear in mind the problem since for the sake of comparison of the Russian documents with the English language documents I will have to do the English translation that didn't exist before.

Reports

Complete and word for word translation from Russian of what is called in English "scientific and technical report" is "report on scientific research work and experimental design". In other words, in a more literary way, we mean "scientific reports on research and development projects". All the organizations in Russia issuing these kinds of reports (an overwhelming majority of which are government funded) are obliged to deliver a free full text copy of scientific and technical report (should we use the term like it is used in Western documents) with a filled-in metadata form to VNTIC in accordance with the Federal Law of the Russian Federation "On the obligatory copy of documents" (the latest amended version of 2002). The status and activities profile of VNTIC is similar to that of National Technical Information Service (NTIS) in the United States with the difference that the former is responsible not only for reports but for dissertations (theses) also.

Twenty years ago, in mid-eighties, every year more than 100 thousand reports arrived at VNTIC. In the nineties this figure dropped 20 times. The reasons for the dramatic fall were the abrupt suspension of scientific research funding so that most of scientific works stopped and the decline of executive discipline in scientific organizations so that

even the completed reports were not always sent to VNTIC. The situation has recently been improving and now the annual arriving reports flow at VNTIC approaches 10 thousand.

Evidently, the centralized collection, processing and archiving of reports suggested that they should have been prepared and presented in a more or less uniform way. This uniformity was regulated by the state, or government standards (abbreviated as GOST) of the USSR and, later on, of the Russian Federation. The standard on scientific report presentation was adopted within the "System of standards on information, librarianship and publishing. The research report. Structure and rules of presentation" – this English translation of the Russian standard's title is given as subtitle in the head of the officially published standard's text. All the rest of the text is in Russian only. Here we see that "the research report" is just another version of translation for "scientific and technical report". Within the system of standards the report standard was assigned index number 7.32 (GOST 7.32 followed by the year of its adoption).

The last Soviet version of the standard was GOST 7.32 – 91. In the most of the main aspects it corresponded the ISO standard "Documentation – Presentation of scientific and technical reports" (ISO 5966/1982) withdrawn in 2000 since several of its requirements became obsolete. The same is true for GOST 7.32 - 91. The next, and still the latest, version of the standard – GOST 7.32 – 2001 [1] – was developed within the framework of the Intergovernmental Council for Standardization, Metrology and Certification of the former Soviet states and prepared by the Intergovernmental Technical Committee on standardization MTK191 "Scientific and technical information, librarianship and publishing". Ten member-states including Republics of Azerbaijan, Armenia, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Tajikistan, Uzbekistan, Russian Federation and Turkmenistan voted for the standard. Since July 1, 2002 GOST 7.32 - 2001 has been officially acting as a state standard of the Russian Federation.

So far, this is the only and most recent standard for reports acting in Russia at least de facto if not de jure since in 2003 the institutions and system of GOSTs were not spared of "reorganization" which is now halfway meaning that the "old" is being ruined but the "new" is not being created because nobody understands what it is. Indeed, don't fix it if it ain't broken. But this is quite another story and we'd better return to GOST 7.32 - 2001.

The nearest Western documents analogous in purpose to GOST 7.32 - 2001 are the European "Guidelines for the production of scientific and technical reports: how to write and distribute grey literature" [2], also known as "Nancy style" recommendations prepared by Paola De Castro and Sandra Salinetti (whom I am obliged for stimulating me to write this paper and for pieces of sound advice) under the aegis of the newly established Grey Literature International Steering Committee (GLISC) in March 2006 and the American national standard ANSI/NISO Z39.18 – 2005 "Scientific and Technical Reports – Preparation, Presentation, and Preservation" [3].

Russian standard GOST 7.32 – 2001 is rather concise (16 pages) and traditional which in this case means a shortcoming since while giving the useful guidelines for report writing and presenting it looks as if it is still oriented on the traditional paper-based publication of reports without the necessary shift of focus concerning the digital representation of documents and corresponding copyright challenges. The standard doesn't differ much from its previous version 7.32 – 91 and withdrawn ISO 5966/1982 standard except for minor additions and cosmetic corrections.

GOST 7.32 – 2001 standard's table of contents is the following:

Foreword
1 Scope of Standard
2 Referenced Standards
3 General Definitions
4 Structural Components of Reports
5 Content Requirements of Report Components Structure
6 Presentation and Display
Appendix A Example of Abstract
Appendix B Examples of Title Pages
Appendix C Example of Authors' List

Foreword gives short formal information about the standard's development and implementation mentioned above.

Part 1 (Scope of Standard) is also very short – just 12 lines in 4 paragraphs and extends the application of the standard to reports on basic and applied scientific research works covering all the branches of science and technology and carried out by scientific research and development organizations, higher educational institutions, industrial enterprises and corporations, joint-stock companies, etc. The last line of the part ("The standard does not apply to scientific research reports in the humanities") is very Russian-specific because in the Russian language the words "science" ("nauka"), "scientific" (nauchny") are applicable to all the branches of knowledge including the humanities and the Arts but not to exact, natural and engineering subjects only.

Part 2 (Referenced Standards) gives the list of 12 rather old Soviet-Russian standards (GOSTs) of "pre-digital" era approved in 1960 – 1995. The referenced standards have to do with documentation and documents, their preparation, presentation and preservation. Although few of the standards may still be valid and useful like, for example, GOST 8.417 – 81 on units of measurement, the most of the standards seem quite obsolete because concern paper-based documentation only.

Part 3 (General Definitions) consists of 10 lines with a short definition of scientific and technical report as official document. It is also stated that besides the final report in the end of research, intermediate reports are possible if they were envisaged by the technical assignment and plan for the research work. The issuing organization is made responsible for the report authenticity, data reliability and conformity to the standard.

Part 4 is the list of 12 Structural Components of Reports:

- **title page;**
- **list of authors;**
- **abstract;**
- table of contents;
- referenced normative documents;
- definitions;
- symbols and abbreviations;
- **introduction;**
- **main part (core) of report;**
- **conclusions;**
- list of references;
- appendices.

The compulsory components are shown in bold type. The conditions for the others are explained in Part 5.

Part 5 contains detailed Content Requirements of Report Components Structure. Its 12 sections (5.1 – 5.12) correspond to 12 structural components given in Part 4. Part 5 is still useful and up-to-date since it describes content, which doesn't depend on the form, that is, whether the report is prepared and presented on paper or digitally.

Part 6 (Presentation and Display) gives the instructions on the form of report organization and is most disappointing since it still deals entirely with paper documents. Computer is mentioned only with printer as a modern typewriter to print the report on paper. No hints on digital media at all. Section 6.1.2 insists that "report shall be printed on one side of A4 sheet of white paper using typewriter (sic!) or computer and printer". Section 6.8.8 allows the equations and formulae to be handwritten in black ink. In short, this approach is completely obsolete, cannot be useful for scientists, engineers and other report authors of today and needs urgent revision.

Appendices A, B, C are self-evident and need no explanation though they are absolutely insufficient as compared to the comprehensive appendices of the standard Z39.18 – 2005.

In general, it becomes clear that the standard GOST 7.32 – 2001 is quite inadequate to the demands of the digital age document turnover. To this effect it is inferior to both the American standard ANSI/NISO Z39.18 – 2005 and the GLISC Guidelines for the production of scientific and technical reports.

Z39.18 – 2005 is the most complete and detailed standard that meets all the modern demands of scientific and technical reports preparation, presentation, preservation, dissemination and archiving. It doesn't reject the paper-based reports but timely shifts the accents onto the digital formats as well as recommendations for multimedia reports. Computers give users much more freedom of choice in the presentation of texts and illustrative material hence it is very important to standardize the digital formats of reports for the ease of their migration, exchange, dissemination and archiving. Z39.18 – 2005 gives comprehensive instructions to this effect. Much attention is paid to metadata and XML descriptions; non-print-specific guidelines are given in separate paragraphs. Unfortunately, nothing of the kind is to be found in the Russian standard.

Though absent from the Russian standards, reports' metadata has been regulated by VNTIC Instructions on the obligatory copy of reports approved on the ministry level. In accordance with the Instructions every full-text report should be accompanied by the so-called "information card" (also called "the secondary document") including the bibliographic data and the abstract of report. This information card is similar to the American Report Documentation Page (shown in Appendix E of Z39.18) or NTIS Bibliographic data sheet – Standard Form 298. The standard form of the VNTIC information card together with the guidelines for completing the form by report authors are given in the Instructions. While American Standard Form 298 is obligatory to be filled-in for federal government agencies and optional for academic or industrial reports the VNTIC information card is obligatory for all the reports no matter what the status of the performing organization is. The information cards arriving at VNTIC are inputted into the retrospective (since 1982) database that serves as electronic catalogue and search aid for VNTIC full-text report collection.

There is another important problem emerged in the digital age that is the problem concerning the intellectual property rights. In fact, there is a complex of problems mainly of ethical nature which did exist in science and grey literature always but in the digital and networking environment grew more acute since the difference between the original and the copy becomes blurred, the accessibility and compilation of texts is facilitated, the plagiarism is easier to perform and harder to trace, etc. Like all the technological innovations the digitalization is ethics-free and enhances the human potentialities in the directions of both good and evil. So, guidelines are necessary to support report authors in the way where they could always stay on the strictly ethical side.

The main advantage of the GLISC Guidelines for the production of scientific and technical reports is the treatment of those problems. The Guidelines are not so detailed as the corresponding ANSI and GOST documents but the GLISC document is unique in addressing the ethics of scientific reporting. In short but informative paragraphs the Guidelines explain sometimes a very delicate distinction between authorship and contributorship, the role of issuing (performing) organization and peer review; typical conflicts of interest are considered and the roots of potential conflicts are outlined. A special attention is paid to copyright considerations and report electronic publishing and archiving. Of course, the Guidelines are not intended to be complete but even the naming and brief commentary of the issues is very helpful.

The report preparation part of the Guidelines is quite traditional but what seems most valuable here is the section on revision editing. There was an illusion that computer-aided text preparation saved the authors from the procedure of editing. Quite timely, the Guidelines urge the authors not to ignore separate procedures of editing. The recommended system of editing includes three levels of revision: rush edit, standard edit and professional edit.

Roaming about the ruins of Soviet science many students and scholars in Russia have lost their ethical orientation and actively use modern digital facilities to report somebody else's results for their own. Following the "saving" rules of the penny-wise and pound-foolish economy many issuing organizations readily refused to edit their

reports giving birth to heaps of ignorant documents. That is why the Guidelines are of special interest to the Russian reader.

Dissertations

There are two post-graduate scientific degrees in Russia: Candidate of Sciences and Doctor of Sciences. Again, the word "sciences" may apply here to all the branches of knowledge, including the humanities. It is possible to draw a parallel between the degrees of Candidate of Sciences and PhD in the USA. "Doctor of Sciences" is superior to "Candidate of Sciences" and seems to have no direct analogue in the West where the title "Dr" may be applied either to PhD or just to a university graduate (like in some European countries). As a rule, one cannot become a Doctor not being a Candidate in Russia.

The would-be Candidates and Doctors are supposed to present a paper with a detailed description of their work and results called "dissertation" (Candidate or Doctoral) and synonymously translated into English as "thesis". (Here again the translator's dilemma: while "thesis" corresponds to the Russian meaning of "dissertation" the latter sounds like in Russian but in some countries may mean just a graduate work). An average (by volume, of course) Candidate dissertation consists of about 150 A4 pages and a Doctoral – about 350 A4 pages.

In general, the dissertation system is much more centralized and unified in Russia than in the West where the scientific degrees are conferred by universities with a great variety in requirements and quality. In Russia dissertations are presented and collected in a unified way with two federal agencies responsible for the procedures: the Higher Attestation Commission (known in Russian in abbreviated form as VAK) and VNTIC. The Russian system seems to have advantages as it supports the unified and rather high quality of dissertations both in form and content, provides for the centralized and complete collection of dissertations, makes it easier and faster to search and have access to dissertations.

The dissertations are approved in two stages. First, after being prepared, they are presented to the so-called Scientific Councils attached to most big scientific organizations like research institutes, universities and other higher educational institutions, etc. (There are several thousands of Scientific Councils in Russia). After a public hearing at the Council's meeting the scientific degree is conferred upon the dissertation's author but not finally. Second, the dissertation along with the Council's decision is sent to VAK where the dissertations are finally approved and the degree is officially conferred upon the scientist.

The legal basis for the VAK activities is the Decision of the Government of the Russian Federation that approves the Regulations of dissertation preparation and presentation. The Regulations are similar to report standards but not so strict and contain the list of obligatory structural components of dissertations, their recommended content and form of presentation. Of course, scientific research is a creative process which can hardly be regulated by any instructions but there is no doubt that the Regulations help to improve the quality of dissertations as coherent documents.

The VNTIC dissertation activities are regulated by the same Federal Law as report collection – "On the obligatory copy of documents". The Law states that the producers of dissertations shall deliver an obligatory free copy of the dissertation to VNTIC within 30 days since its presentation. Like in case of reports there are VNTIC Instructions on the obligatory copy of dissertations approved on the ministry level. The Instructions are coordinated with the VAK regulations and the Federal Law. In accordance with the Instructions every full-text dissertation should be submitted to VNTIC together with the secondary document – the dissertation information card containing the bibliographic description and the abstract of the dissertation. The cards data are recorded in the VNTIC retrospective database to facilitate dissertation search and retrieval.

It is interesting to note that in the nineties the annual amount of presented dissertations didn't suffer the same dramatic drop as for reports. In Soviet times about 25 thousand dissertations arrived at VNTIC every year including approximately 20

thousand Candidate and 5 thousand Doctoral dissertations. The same figure with minor random fluctuations has been valid up to now
which in fact means nearly 30% increase because in the Soviet Union around 70% of dissertations originated from the Russian Federation and 30% came from the other Soviet republics where the dissertations aren't sent from to VNTIC now. Moreover, the dissertation flow to VNTIC has lately even grown to annual 30,000.

The phenomenon, good as it is, may have several reasons. Two possible reasons should be mentioned – a positive one and a negative one. The positive reason means that the individual creative activity like scientific research and dissertation writing (which is, by definition, an individual work unlike the report preparation that is collective by nature) cannot be stopped even by the political and economic chaos and financial shortage. The negative reason is the growing amount of poor quality dissertations to say nothing of the shameful and illegal business of dissertations compiling for money to orders of dishonest people. The scale of this business in Russia is really troubling and means a severe ethical degradation of some part of scientific and academic community.

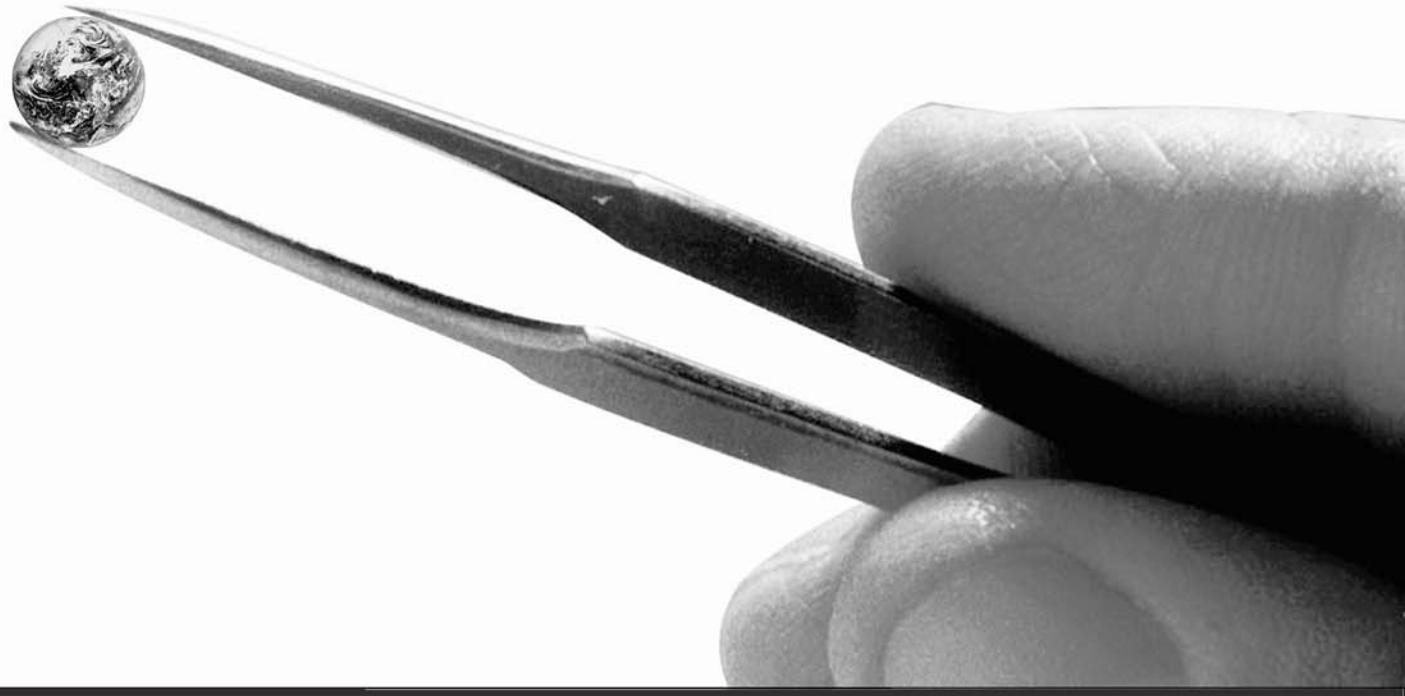
Concluding remarks

Scientific and technical grey literature development in Russia has been supported by a reliable and effective system founded in the sixties of the last century. The system still works but was not updated for fifteen years. Now its normative basis needs urgent revision to reflect the digital document preparation and turnover. The recent Western documents – the American Z39.18 – 2005 and the European GLISC Guidelines – in many respects meet the new requirements and should be carefully studied so that new issues of GOST 7.32, VAK Regulations and VNTIC Instructions could correspond to the digital realities of today.

References

1. GOST 7.32 – 2001. System of standards on information, librarianship and publishing. The research report. Structure and rules of presentation. Intergovernmental standard. Official publication. Minsk, 2001. (In Russian).
2. Guidelines for the production of scientific and technical reports: how to write and distribute grey literature. Version 1.0. GLISC. March 2006. www.glisc.info.
3. ANSI/NISO Z39.18 – 2005. Scientific and Technical Reports – Preparation, Presentation, and Preservation. NISO Press, Bethesda, Maryland, USA. 2005. www.niso.org.

https://doi.org/10.26069/greynet-2026-000.606-tgj



https://doi.org/10.26069/greynet-2026-000.606-tgj

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

Joachim Schöpfel, Christiane Stock, and Nathalie Henrot *(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.

* First published in the GL8 Conference Proceedings, February 2007

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.

https://doi.org/10.26069/greynet-2026-000.606-tgj

Scientists produce and use grey literature, but are they aware of the implications of doing so? *

Paola De Castro *(Italy)*

Abstract
Grey literature includes documents produced by academia, government, and industry outside the commercial circuit and it is mainly represented by technical reports. It is an important primary source of information that is now available through the internet. Because grey literature is not generally subject to the peer review process, it is fundamental that authors and issuing organizations become aware of the implications of the uncontrolled diffusion of specific information. For this reason, the application of ad hoc guidelines ("Nancy Style") for the production of grey literature is highly recommended.

To become a scientist, you must get a degree and produce a dissertation or a doctorial thesis; to report the results of research supported by a grant, you must draw up an activity report or a progress report; to take part in a conference, you must prepare an abstract or a paper to be included in the conference abstract book or in the conference proceedings; to reach consensus on guidelines or regulations on specific topics, you produce a working paper to be discussed within small expert groups; to report extended data on a research project, to draw up operational procedures, or to refer to clinical trials, you write a technical report; to receive comments on a paper before submitting it to a journal, you circulate it among colleagues as a pre-print; to get up-to-date information on a specific issue, you may need to access a governmental report; to perform a correct meta-analysis, you must look at all sources of available information, in both indexed and non-indexed journals; to better understand a document written in a language that is not your own, you may read an unofficial translation; to gain information about the technical characteristics of new equipment, you may require technical or commercial documentation; and to spread general information to selected targets, you produce information leaflets.
These are just a few examples of those precious primary information sources that are produced outside the commercial circuit and that fall under the general term "grey literature", defined as "Information produced on all levels of government, academia, business and industry in electronic and print formats not controlled by commercial publishing, i.e. where publishing is not the primary activity" (Cassell 2005).

Some historical aspects
Some writers consider that grey literature existed during Roman times, and that the drawings and notes of ancient scientists could be regardedas ante literam grey literature. Indeed, the first technical reports — the most typical example of grey literature today — were produced at the beginning of the last century in a military environment, first in Great Britain in 1909 (Research Memoranda of the Aeronautics Research Council) and soon after, in 1915, by the US National Advisory Committee for Aeronautics.
After the Second World War, technical reports were a useful means of communicating information at a time when journal publication was rather slow and information needed to be disseminated rapidly — the informal channel met these requirements. From that time on, there was a rapid development of report production, mainly in the scientific field, where speed of information transfer was more important than in other sectors.
During the last century, however, it was very difficult to retrieve grey literature documents because there were only ever a few copies available and because references to these documents were not included in bibliographic databases. Furthermore, these documents were often produced without the necessary identification items. Indeed, some of them even circulated without the names of the authors or issuing organization, information that represents the "bare minimum" for identifying and obtaining an unpublished document.

* First published in *EASE*, European Science Editing, Vol. 32 (4) November 2006

In 1985, the SIGLE (System for Information on Grey Literature in Europe) bibliographic database was established, with the support of the European Community, to collect and disseminate information about grey literature produced in Europe. Until the use of the internet became widespread, this multi-sectorial database represented a useful reference tool, but in 2005 it was officially abandoned as it no longer fulfilled current information needs.
Today, grey literature is available through the internet, mostly without restriction, and is also included in the catalogues of major libraries (e.g. the catalogue of the US National Library of Medicine). Research in public health relies significantly on grey literature (Alberani et al. 1990, Gray Bedford 1998), but retrieval facilities are not yet sufficiently developed compared with those of peer-reviewed articles. The New York Academy of Medicine's Grey Literature Report (bimonthly online publication alerting readers to new grey literature in public health) represents a good attempt to disseminate important research output (available from: www.nyam.org/library/greyreport.shtml).

Characteristics of grey literature
Researchers today widely use and produce grey literature, but they are often unaware of the implications behind such non-conventional documents (Alberani & De Castro 2001). In the grey literature, the ratio between production costs and the intrinsic value of the document has always been inverted — funds granted to carry out research activities reported in the document far exceed the costs of its production and distribution.
Until a few decades ago, the adjective "grey", with its negative implications, was a perfect description of particular research output that was very poor in its formal attributes (shape, structure, editorial standards, etc.) and was used as opposed to the terms "white" or open literature (journal articles and books) and "black" or classified literature that was not to be disseminated at all. However, the information contained in the grey literature has always been very precious and unique because it was not to be found elsewhere: even if journals sometimes published originally "grey" documents, they did not include all details of the research contained in a report (i.e. descriptions of equipment, procedures, raw data, tables, graphs, maps, etc.), due to limitations of space or because such specific information might not be of interest to a wider audience. Furthermore, despite the many benefits of the peer review process of the most prestigious journals, only statistically significant findings are generally published in the open literature and this often inflates the perceived importance of the reported results. Raising awareness of the hidden value of grey literature may represent an important step in gather ingprecious and unbiased information (Banks 2006).
Traditionally, grey literature was produced in-house with a limited numbers of copies for specific aims and, generally, without following proper editorial standards. Because manuscripts were circulated as written by the authors, without editorial support, sometimes even the basic elements of structure and readability were missing and the bibliographic elements necessary to allow their identification were often lacking. This made it very difficult to retrieve documents outside the "invisible colleges" to which they were addressed (committee members, expert groups, decision makers, etc.).
In the 1970s and 1980s — when the internet was not yet widespread — such informal and rapid circulation of documents was particularly welcome, firstly and mostly by physicists, who are recognized as the pioneers in pre-print (now e-print) production as a way of obtaining colleagues' approval or advice before submitting an article for publication in specialized journals. Grey literature then reached only the target readers and this was a safeguard against any possible misinterpretation or misuse of "unpublished" documents. Copyright laws and ethical considerations could be disregarded.
Today, thanks to the internet, the limited run of documents produced in-house is no longer a hindrance to the dissemination of information, and as a result of the widespread use of new information communication technologies the shabby look of grey literature has rapidly changed.

Open access and the new responsibilities to grey literature authors and producers
Now that access to full-text grey literature is much easier, also thanks to the development of institutional repositories and digital archives, both authors and producers have new responsibilities. In fact, whereas in the past grey literature was addressed only to readers who were directly involved in the issues described, today it is no longer possible to have complete control of the target readership, and the

general public may also access online grey literature, with the consequent risk of misinterpretation. Malevolent readers could misuse specific instructions intentionally addressed only to technical or medical staff (De Castro & Napolitani Cheyne 2006).
In this context, it is essential to guarantee that the grey literature available on the internet meets basic standards of editorial quality and ethical principles. A new challenge therefore arises in making authors and issuing organizations aware of their responsibilities. The development of open access and institutional repositories offers new possibilities for information storage and retrieval.
Until the end of the 1970s, there was little concern in Europe about the presentation of grey literature. In 1982, the ISO standard 5966 on the presentation of technical reports (International Organization for Standardization 1982) was issued as a first step towards better report production. Today, most authors can autonomously produce and disseminate a document, even without editorial support. In most cases, the layout of "unpublished" material is no longer so "grey", and the difference between grey or white literature might not be immediately clear. In this evolving scenario, ISO 5966 — which was very useful until a few decades ago — could no longer meet the requirements of information technology and it was withdrawn in 2000.
In the internet era, due consideration must be given to ethical principles when issuing grey literature, as well as to the possibility of applying a review (or peer review) process to guarantee quality. Authors of grey literature should be aware of editorial rules even more than authors of documents to be published in the open literature because they do not benefit from the referees' or editors' contribution during the editorial process, which provides added value to the original papers.
Recognizing the value of "Vancouver style" (Uniform requirements for manuscripts submitted to biomedical journals, produced by the International Committee of Medical Journal Editors [ICMJE] — available fromwww.icmje.org) for authors and editors of journal articles, and the lack of freely available and updated guidelines for the production of technical reports, some producers of institutional reports felt the need to develop an ad hoc style for grey literature, to be used as a reference tool for its production and dissemination (De Castro & Salinetti 2006a). To this end, a proposal to draw up updated guidelines for grey literature producers was presented at the 7th International Conference on Grey Literature held in Nancy, France, on 5–6 December 2005. The so-called "Nancy style" for the production of technical reports was developed on the basis of this proposal and was issued in April 2006.

Guidelines for the production of grey literature: Nancy style

"Nancy Style" is the informal name given to the Guidelines for the production of scientific and technical reports: how to write and distribute grey literature, which are freely available in English, French, and Italian from www.glisc.info. The Guidelines were created to help the authors and producers of grey literature to write and distribute accurate, clear, and easily accessible reports. As mentioned, the Guidelines were adapted from the "Vancouver style" and the ISO standard 5966/1982 and cover:

- ethical considerations, including the responsibilities of authors, contributors and issuing organizations (playing the role of editors), conflicts of interest, peer review, security concerns, etc.
- publishing and editorial issues, including copyright, electronic publishing, institutional repositories, etc.
- recommended report structure, including the compulsory elements of a report, sections, formats, styles, illustrations, references, appendices, indexes, etc., and principles of revision editing.

A correct report structure is essential to guarantee readability and ease of use. A well-organized document will also be easily converted into XML, to allow advanced search facilities in specific parts of the document, such as the introduction, conclusion, and citations.
Institutions producing grey literature should be encouraged to draw up instructions to authors to guarantee a standard structure that is consistent with institutional policy and editorial and ethical considerations. They are also advised to provide adequate training for authors unaware of editorial standards.
Since grey literature is not generally peer reviewed, it is essential that authors should understand the importance of careful revision of their texts before their distribution, to improve correctness, quality, and style. Detailed information on "Nancy Style" and the new responsibilities of authors and producers of grey literature is reported in a recently published chapter of the Science Editors Handbook (De Castro & Salinetti 2006b).

Final considerations

The newborn Nancy style, like the Vancouver style, could develop into a de facto standard, representing uniform requirements for the production of technical reports produced by an international group — the Grey Literature International Steering Committee, GLISC (for information see www.glisc.info) — in an effort to combat ignorance of the best editorial practices and to facilitate integration, cooperation, and standardization. The promotion of Nancy Style by EASE could be part of a broader programme to raise awareness of European science editing (Hunt 2006). Even if grey literature is traditionally considered as a sui generis editorial product, it is nevertheless a written documentation of research output that deserves due consideration.

Grey literature is not associated with any prestige or career implications, but training inexperienced authors to produce grey material may represent the first step to publication at a higher level. Tutoring in scientific writing and data presentation also helps to improve research methodology. Experience, gained over years of writing courses in the biomedical field, shows that when a scientist understands the reasons why editorial rules and standards must be applied, his or her publication output will improve rapidly, even thoughthe challenge of having an article accepted by mainstream journals still represents the ultimate aim of all researchers.

References

Alberani V, De Castro P. 2001. Grey literature from the York Seminar (UK) of 1978 to the year 2000. INSPEL 35(2)36-47. Available from: www.ifla.org/VII/d2/inspel/01-4alvi.pdf; accessed 7/4/06.

Alberani V, De Castro P, Rossi AM. 1990. The use of grey literature in health sciences: a preliminary survey. Bulletin of the Medical Library Association 78(4): 358-363.

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

Cassell KA. 2005. Report on the 6th International Conference on Grey Literature, New York. Collection Building 24(2):70–71.

De Castro P, Salinetti S. 2006a. "Uniform Requirements" for grey literature: proposal for the adoption of "Nancy style". Publishing Research Quarterly 22(1):12-17.

De Castro P, Salinetti S. 2006b. Writing and issuing grey literature: old and new responsibilities. In: Maissonneuve H, Enckell PH, Polderman AKS, Thapa R, Vekony M (Ed.). Science editors' handbook. Old Woking: European Association of Science Editors.

De Castro P, Napolitani Cheyne F. 2006. Small science journals: stay alert for potentially dangerous information. European Science Editing 32(1):11–13.

Gray Bradford H. 1998. Sources used in health policy research and implications for information retrieval systems. Journal of Urban Health 75(4):842–852.

Hunt R. 2006. Communicating European research 2005. European Science Editing 32 (1):16–18.

International Organization for Standardization. 1982. ISO 5966: Presentation of scientific and technical reports. Geneva: ISO.

https://doi.org/10.26069/greynet-2026-000.606-tgj

Ninth International Conference on Grey Literature

Grey Foundations in Information Landscape

GL9

Conference Announcement

The **Ninth International Conference on Grey Literature** seeks to map the infrastructure in which grey literature is embedded. This concerted drill in the field of information also stands to further a framework for shared understanding.

The title of the conference **'Grey Foundations in Information Landscape'** encompasses five main themes:

- Newfound Grey Resources in Electronic and Print Formats
- Tools for Publishing, Archiving, and Accessing Grey Literature
- Uses and Applications of Grey Literature in the Social and Human Sciences
- Grey Inroads to a Multicultural and Multiethnic Urban Village
- Grey Literature in Central and Eastern Europe

GL9 will survey the information landscape by exploiting newfound and existing grey resources, by utilizing tools developed in part or whole for processing grey literature, and by clearly demonstrating uses and applications for research and policy driven settings.

The conference venue will provide information professionals with a variety of platforms for presenting and communicating results. Plenary sessions will be held in the Antwerp Council Hall. In the meeting area adjacent to the lobby, an 'Information Walk-Thru' will accommodate Product and Service Reviews as well as Poster presentations. And, the Birds-of-a-Feather lunches will offer flexible seating arrangements for special interests groups.

GL9 Program and Conference Bureau

TextRelease

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

House of the Province, Antwerp, Belgium, 10-11 December 2007
https://doi.org/10.26069/greynet-2026-000.606-tgj

Ninth International Conference on Grey Literature

Grey Foundations in Information Landscape

GL9

Call for Papers

Guidelines for Submission of Abstracts

For:

Authors, Publishers, Librarians, Web Editors, Researchers, Policy Makers, Information Managers, Brokers and Vendors, Information Specialists, Intermediaries, Information Technicians, Information Professionals, Journalists, and Academia

GUIDELINES FOR THE SUBMISSION OF ABSTRACTS:

Participants who seek to present a paper at GL9 are invited to submit an English abstract between 400-500 words. The abstract should deal with the problem/goal, the research method/procedure, an indication of costs related to the project, as well as the anticipated results/conclusions of the research. The abstract should include the title of the paper, names of the author(s), and full address information.

TITLE OF YOUR PAPER

ADDRESS INFORMATION

AUTHOR'S NAME(S)	**T**ELEPHONE
ORGANIZATION(S)	**F**AX
ADDRESS/P.O. BOX	**E**MAIL
POSTAL CODE - CITY - COUNTRY	**U**RL

NB *Abstracts are the only tangible source, which will allow the Program Committee to guarantee content and balance in the plenary sessions. Every effort should be made to reflect the content of your work in the abstract submitted. Abstracts not in compliance with the Guidelines may be returned to the author for revision.*

INDICATE THE SESSION THEME MOST SUITED TO YOUR ABSTRACT:

- NEWFOUND GREY RESOURCES IN ELECTRONIC AND PRINT FORMATS ☐
- TOOLS FOR PUBLISHING, ARCHIVING, AND ACCESSING GREY LITERATURE ☐
- USES AND APPLICATIONS OF GREY LITERATURE IN THE SOCIAL AND HUMAN SCIENCES ☐
- GREY INROADS TO A MULTICULTURAL AND MULTIETHNIC URBAN VILLAGE ☐
- GREY LITERATURE IN CENTRAL AND EASTERN EUROPE ☐

DUE DATE AND FORMAT USED FOR SUBMISSION:

The abstract must be emailed on or before May 1st 2007 in MS Word. The author will receive written verification upon its receipt. The GL9 Program Committee will then use these abstracts to finalize the Conference Program.

FURTHER PROCEDURE:

After the GL9 Program Committee meets in June, those who submitted abstracts will receive notification of their place on the conference program. This will also be accompanied by guidelines for the submission of the full-text of conference papers and accompanying PowerPoint presentations.

GL9 Program and Conference Bureau

TextRelease

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

https://doi.org/10.26069/greynet-2026-000.606-tgj

About the Authors

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. Marcus 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

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

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

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

Leonid P. Pavlov graduated from Moscow Physical-Engineering Institute, Dipl. Eng. in computer systems. He is a Candidate of Sciences in informatics; and since 1976 is employed with the Scientific and Technical Information Centre of Russia (VNTIC) as Deputy Director. Main works in information systems, scientific and technical information, and grey literature.
Email: pavlov@vntic.org.ru

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

https://doi.org/10.26069/greynet-2026-000.606-tgj

Notes for Contributors

Non-Exclusive Rights Agreement

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

Submission Requirements

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

REFERENCE GUIDELINE

General

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

Specific

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

Examples:

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

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

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

Review Process

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

Journal Publication and Article Deposit

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

https://doi.org/10.26069/greynet-2026-000.606-tgj
TGJ Volume 3, Number 1 2007

An International Journal on Grey Literature

'GREY STANDARDS IN TRANSITION AND USE'

SPRING 2007 – TGJ VOLUME 3, NUMBER 1

Grey Literature - Taxonomies and Structures for Collection Development 7
Julia Gelfand *(United States)*

Metadata-based analysis to improve clinical trial exchange 17
Daniela Luzi, Fabrizio L. Ricci *(Italy)*, and Luca Dan Serbanati *(Romania)*

Implications of Copyright Evolution for the Future of Scholarly Communication and Grey Literature 31
Marcus A. Banks *(United States)* and Cees de Blaaij *(Netherlands)*

Legal Foundations of the Scientific and Technical Grey Literature Development in Russia 37
Leonid P. Pavlov *(Russia)*

From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource Migration in European Context
Joachim Schöpfel, Christiane Stock, and Nathalie Henrot *(France)* 45

Scientists produce and use grey literature, but are they aware of the implications of doing so?
Paola De Castro *(Italy)* 52

Colophon 2
Editor's Note 5
Advertisements
- EBSCO Information Services 4
- ERIC-NLE, National Library of Education 6
- GESIS, German Social Science Infrastructure Services 36
- INIST-CNRS, Institute for Scientific and Technical Information 44

GL9 Announcement and Call-for-Papers 56
About the Authors 58
Notes for Contributors 59