

Grey Matters in the World of Networked Information

# FIFTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

KNAW CONFERENCE CENTER  
AMSTERDAM, THE NETHERLANDS

DECEMBER 4-5, 2003



Celebrating GL Conferences 1993-2003

Conference Proceedings

## **Sponsors**

**INIST-CNRS**

*Institut de l'Information Scientifique et Technique;  
Centre National de Recherche Scientifique, France*

**City of Amsterdam**

*Department of Communication, External Affairs*

## **Program Committee**

**EAGLE**

*European Association for Grey Literature Exploitation*

**GreyNet**

*Grey Literature Network Service, Netherlands*

**INIST-CNRS**

*Institut de l'Information Scientifique et Technique, France*

**JST**

*Japan Science and Technology Agency*

**OSTI-DOE**

*Office of Scientific & Technical Information, U.S. Department of Energy*

**UiB**

*University of Bergen, Norway*

**UCI**

*University of California, Irvine, United States*

**UTwente**

*University of Twente, Netherlands*

**TextRelease**

[www.textrelease.com](http://www.textrelease.com)

GL5 Program and Conference Bureau

Beysterveld 251, 1083 KE Amsterdam, The Netherlands

Tel/Fax: +31-(0) 20-672.1217 Email: [info@textrelease.com](mailto:info@textrelease.com)

## **CIP**

### **GL5 Conference Proceedings**

Fifth International Conference on Grey Literature : Grey Matters in the World of Networked Information, 4-5 December 2003 / comp. by D.J. Farace and J. Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, January 2004. VI, 162 p. : ill. ; 30 cm. - Author Index. - (GL-conference series, ISSN 1386-2316 ; No. 5)

The conference proceedings contain the full text of some twenty-one papers presented during the Opening, Plenary, and Breakout Sessions. The print copy has been standardized in font, size, and format for general layout and appearance. Included is an Author index and full address information.

## Foreword

**A**fter a four-year interval, the fifth in the series of international conferences on grey literature took place in Amsterdam on 4-5 December 2003. Sixteen countries worldwide were represented at GL5, from Finland to South Africa and from Japan to the United States. Twenty-four conference papers were presented in sessions geared to research, models and strategies, as well as the technical know-how of search engines and roadmaps for GL systems and services.

This Conference Proceedings provides the papers from GL5 - the goal of which is reflected in the conference title: "Grey Matters in the World of Networked Information". Since its redefinition at the Luxembourg convention in 1997, grey literature sought juxtaposition to commercial publishing, where grey is not controlled. However, while the exploitation of grey literature resources are not primarily for economic gain, they do augment the knowledge and information bases in specialized fields, empower decision making in local and national government, enable forecasts for business and industry, and impact public opinion of net-citizens.

At GL5, grey literature now seeks juxtaposition to networked information, where networked can be interpreted differently. On the one hand, it can refer to the electronic environment in which it is produced, processed, and distributed; or, on the other hand, to the organisational structure and system in which it is produced, processed, and distributed. Grey literature must adhere to both interpretations. The guarantee GL provided in the pre-Internet era still applies. In brief, responsible and recognised organisations in government, academics, business and industry stand behind the content of the information. Within each of these sectors there is an established (peer) review process. Further, the document and its related documentary information can be identified by its type e.g. working paper, preprint, thesis, etc., by the standards used in its production and processing, and by its incorporation in a collection, where availability through loan and/or document delivery is provided by libraries, depots, and information centres. Today, an enormous volume of information found on Internet does not offer these guarantees, and is not grey. In fact, this is the foremost challenge to the global grey literature community, enabling net-users to make the distinction.

While not all of the questions raised were answered, the direction and momentum of this international Conference Series has never been more manifest. I invite you to read this selection of papers; and on behalf of the authors I welcome your thoughts and ideas for GL6, where work on grey is in progress!

## Conference Sponsors



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



**City of Amsterdam**  
Host of the GL5 Inaugural Reception

## Program Committee Members



**Prof. Dr. Hans E. Roosendaal, *Chair***  
University of Twente



**Anne Asserson**  
University of Bergen



**Takaharu Atago**  
Japan Science and Technology Agency



**Deborah E. Cutler**  
Office of Scientific and Technical Information;  
U.S. Department of Energy



**Dr. Dominic J. Farace**  
Grey Literature Network Service;  
TextRelease



**Julia Gelfand**  
University of California, Irvine



**Dr. Joachim Schöpfel**  
European Association for Grey Literature Exploitation



**Christiane Stock**  
Institut de l'Information Scientifique et Technique  
Centre National de Recherche Scientifique

## Contents

Foreword	iii
Sponsors and Program Committee Members	iv
Conference Program	vi
Opening Session <i>Grey Matters in the World of Networked Information</i>	1-12
Session One <i>Models for Academic Grey, Part I: Specific Approaches</i>	13-31
Session Two <i>Research is Grey Dependent</i>	32-63
Session Three <i>The Economy of Grey</i>	64-74
Session Four <i>Strategies for Academic Grey, Part II: General Approaches</i>	75-105
Session Five <i>Search Engines are Growing Grey</i>	106-116
Session Six <i>Roadmap of Grey Literature Systems and Services</i>	117-131
Breakout Session <i>Alternative Issues in Grey Literature</i>	132-158
<b>APPENDICES</b>	
List of Participating Organisations	159-160
Index to Authors	161
Publication Order Form	162

**Opening Session**

H.M. Artus - Old WWine in New Bottles? Developments in electronic information and communication: Structural change and functional inertia	1
D.J. Farace, J. Frantzen - Four Winds on the Grey Landscape: A Review of Four Information Professionals, Their Work and Impact on the Field of GL	10

**Session One - Models for Academic Grey, Part I: Specific Approaches**

G.E. Siegel - Capturing Academic Grey Literature: Starting at Home	13
J. Gelfand - The Ubiquity of Grey Literature in a Connected Content Context	21
M. Claerebout - Grisemine, a digital library of grey university literature	27

**Session Two - Research is Grey Dependent**

P. De Castro, S. Salinetti - Quality of grey literature in the open access era: privilege and responsibility	32
D. Luzi, M. Castriotta, R. Di Cesare, L. Libutti, M. Manco - The Communication Flow of Research Projects Results	40
B.H. MacDonald, R.E. Cordes, and P.G. Wells - Grey Literature in the Life of GESAMP, an International Marine Scientific Advisory Body	50

**Session Three - The Economy of Grey**

L.P. Pavlov - The Commercialization of Research Findings Documented in Grey Literature	64
W. Roem, M. Minderhoud - Collection development in support of a global information network: A case study of LEISA, Low External Input and Sustainable Agriculture	69

**Session Four - Strategies for Academic Grey, Part II: General Approaches**

H.E. Roosendaal - The Information Market for Research and Higher Education: How to integrate all relevant information in a network of repositories?	75
A. Kairamo - The moving border of tacit and explicit knowledge in e-Learning: Use and production of information and knowledge in technical university education, case study	87
C. Stock, J. Schöpfel - Grey literature in an open context: From certainty to new challenges	94

**Session Five - Search Engines are Growing Grey**

M. Vesely, J. Le Meur, T. Baron, T. Simko - CERN Document Server: Document Management System for Grey Literature in Networked Environment	106
C. de Blaaij - Grey literature from invisibility to visibility: The quest for grey content in the domain of the invisible web	112

**Session Six - Roadmap of Grey Literature Systems and Services**

C. O'Dell, D. Dallman, M. Vesely, J. Vigen - 50 Years of Experience in Making Grey Literature Available: Matching the Expectations of the Particle Physics Community	117
I. Pitoni, D. Macri - GL systems and services in the specific fields of vocational training and labor policies in Italy: the ISFOL Case	124

**Breakout Session - Alternative Issues in Grey Literature**

H. von Hofe - Emergent Recognition: An Alternative Perspective on the Grey Literature of an American Public High School, 1985-2001	132
W. Theaux - Body and Literature: Study of the Grey Literature of a Health Service Starting from a Reflection on Health	140
I. Pitoni, D. Macri - The ISFOL Specialised Documentation Centre: an important tool for operational research	147
Maria do Rosário Guimarães Almeida [et al] - Management of GL for the construction of a database for the University of Maranhão's - UFMA and the State University of Maranhão - UEMA	152

## Old WWWine in New Bottles? Developments in electronic information and communication: structural change & and functional inertia

Dr. Helmut M. Artus  
Information Centre for the Social Sciences (IZ)

### *Abstract:*

To understand the meaning of grey literature in the internet age, it could be helpful to have a look at the whole system of scientific information and communication and its changes at the time being, not only at its informal ('grey') sector.

- *Part 1* of this paper presents empirical data on electronic journals ('e-journals') in the Social Sciences in Germany and the differences between commercial and non-commercial publishers. Further information concern different types of electronic publications - 'grey' and 'white' - and innovative developments in electronic information and communication *via* the internet. This is exemplified by a network of R&D projects in Germany under the general name *vascoda*. The project's vision is an integration of searching information on literature in distributed sources and receiving the full texts immediately, preferably by download. The Social Sciences are represented by the *infoconnex* project which includes grey literature; *infoconnex* is part of the *vascoda*-network.
- *Part 2* draws some (theoretical) consequences for our understanding of grey literature and its future. As to the *technical facilities*, things have widely changed (although printed grey literature is not completely out of fashion), but with regard to the *economic structures* and *social functions* of grey literature we can confirm that only little has changed - if anything at all.

### Introduction

It was in late 1994, that my institute offered its first page in the internet and proudly presented a so-called *gopher* which was the most advanced technical device of information technology at that time. A *gopher* is a kind of squirrel, i. e. a small red haired animal with a bushy tail climbing up and down the trees, and that is why it gave its name to a structure of internet files arranged like branches of a tree.

This advanced technology, however, was at the same time the reason for failure since hardly any of our customers - which means: of *social scientists in Germany, Austria, and Switzerland* - had an internet connection at his disposal at that time. Most of them did not even know what the internet was and what it was good for. It still took years until the social scientific community discovered the internet which meanwhile had switched over to the *world wide web* structure (WWW) as organizing principle or *hypertext transfer protocol (http)*. The situation in the Social Sciences was not much better when we made a relaunch of our internet pages, now in the *world wide web* in late 1996.

Now, only seven years later, the internet is a self-evident tool for information and communication even in the Social Sciences.

The internet provides all possibilities we have always been dreaming of. All we need is a computer, internet access and a homepage with a URL - the *unified resource locator*, which is the internet address of a site. Then, we can be all in one: the *author*, the *board of referees* deciding whether an article of ours is accepted for publication or refused, the *printer*, the *publisher* and the *book seller*. And it does not cost a penny more to distribute our articles worldwide.

If this happens - what is going to happen with grey literature? Grey literature can only be defined in its relation to formally published or 'white' literature. This white literature, however,

we could expect, this white literature is going to perish in the long run - and grey literature with it. What remains, is only "literature" without any further classification as white or grey, formally or informally published and so on - just *literature*.

The internet has changed all our information and communication habits. Why shouldn't it change our publishing habits either? !

### Chapter 1: Empirical Findings

In the following text, I try to test this assumption. Chapters 1 and 2 are dedicated to empirical findings as well as to well-founded yet not quantifiable observations from different R&D projects.<sup>1</sup> Chapter three tries to give a theoretical explication of these findings.

### *The Case of Internet Journals*

Let us start with scientific journals. My data refer to 101 internet journals in the Social Sciences, mostly from the German speaking countries. As far as we know, this is nearly a full sample.<sup>2</sup> The short information in the database was completed by further information found on the journals' homepages.

From these 101 journals, 42 are commercial and 59 are not.

**Table 1: e-Journals According to Type of Publisher**

<i>Type of Publisher</i>	<i>abs.</i>
Commercial	42
Non-commercial.	59
<b>Total:</b>	<b>101</b>

The non-commercial publishers consist of scientific institutes (34) - mostly from universities -, public bodies or authorities (12) and others (13).

At first glance, it seems that this is a splendid corroboration of the assumption that grey literature is on its way and formally published white literature is on its retreat. Only forty percent of internet journals, one could argue, still come from commercial publishers.

These 101 e-journals published electronically or at least having an electronical version represent nearly 30 percent of all about 350 titles regularly indexed by the *Information Centre*. So, this is also a considerable part of the total journal production.

But this is much too simple. Before daring such a conclusion, we have to answer several other questions, e. g.:

- Are journals published by commercial publishers really liable to charge and journals from non-commercial publishers really free of charge? According to our assumption there should be a considerable trend to free distribution even with commercial publishers.

<sup>1</sup> In a certain sense, this is the empirical test of the prognosis I made some years ago:  
*Helmut M. Artus: SMAISMRMILMEPOETALEUMIBUNENUGTTAUIRAS. The internet & the socio-structural change of informal scientific communication, in: New Frontiers in Grey Literature. Fourth International Conference on Grey Literature, Washington D. C., 4-5 October 1999, p. 12-24*

<sup>2</sup> The information are derived from the database "Zeitschriften" (=Journals) of the Information Centre for the Social Sciences. For information (IZ) on the database and its contents:  
<http://www.gesis.org/Information/Zeitschriften/index.htm>; for information search in the database itself:  
<http://193.175.239.137/zsdb/servlet/zsdb.ZSFormular>. - All data as of October 2003.



- Have there been transitions from commercial to non-commercial publishers and publishing?
- And eventually: Is the distribution between commercial and non-commercial publishers really induced by the internet - or is it just a 'projection' of an old situation already existing in times of print or photo copier?

**Table 2: e-Journals: Free or not free?**

<i>Publisher</i>	<i>Internet Version</i>			
	<i>free of charge</i>		<i>liable to charge</i>	
	<i>abs.</i>	<i>%</i>	<i>abs.</i>	<i>%</i>
Commercial	6	14	36	86
Non-commercial	50	85	9	15

These data are quite nice because the percentages for commercial and non-commercial publishers are nearly exact mirror-images: 14 percent of e-journals from commercial publisher are free - and 15 percent from non-commercial publishers are liable to charge. Accordingly, 86 percent from commercial publishers are liable to charge, and 85 percent from non-commercial publishers are free. The exact differences are less than 1 percent point.

Fourteen percent of *free* e-journals from *commercial* publishers, however, could be regarded as a relatively high percentage indicating a trend towards a softening of commercial and professional standards. But I think, that this is not pertinent. All these 6 e-journals are exclusively accessible via the internet, i.e. they do not have a printed version.

In fact, none of these six e-journals is a scientific journal in a strict sense, but rather a journal combining short articles with more or less relevant information concerning a scientific discipline or a subject matter.

So I think this is not the case of commercial publishers making concessions to a free and non-commercial internet, but it rather seems, that the commercial publishers did not let the opportunity slip to raise their attraction by expanding their free internet offer, particularly because digital publication is less expensive.

This interpretation is supported by the data concerning the print versions of the e-journals. Only one e-journal from a commercial publisher has no print version. All other 27 print versions have their price - and it is a real, commercially calculated price, not a symbolic one. (For 14 e-journals from commercial publishers, no information was available.)

This shows convincingly that the commercial publishers have by no means given up their claims but have only expanded to a new form of market, i. e. the digital market of the internet. For them, the internet is not a realm of freedom, but it is a big commercial market, which also allows a reduction of costs for production and distribution. This reduction of costs, I think, is the main reason for non-commercial publishers to offer their journals in most cases only in an electronic version.

**Table 3: Print Versions of e-Journals: Free or not free?**

<i>Publisher</i>	<i>Print Version</i>		
	<i>NO</i>	<i>free of charge</i>	<i>liable to charge</i>
Commercial	1	-	27
Non-commercial	22	9	11

In 14 cases (commercial) and in 17 cases (non-commercial) no clear or reliable information was available, be it on prices, be it on the existence of a print version.

Before finishing the topic of e-journals, we should still ask for transitions from commercial to non-commercial publishing (table 4). To be honest, my data only show *differences*, not *directions* or *transitions*. But if seen in the whole context, I think, an interpretation in this sense is admissible.

The first case with the heading “*No Difference between Internet and Print*” is clearly predominant: 83 percent vs. only 17 percent with a difference. As to commercial publishers, in nearly all cases (24 of 27) there is no difference - which means that in all these cases you have to pay - no matter which version or medium you prefer:

**Table 4: Differences Between Internet and Print Version\*)**

Type of Publisher	No Difference between Internet and Print		Difference between Internet and Print		Total
	Free of charge	liable to charge	Internet free of charge	Print version free of charge	
Commercial	-	24	3	-	27
Non-commercial	9	6	5	-	20
<b>Total:</b>	<b>9</b>	<b>30</b>	<b>8</b>	<b>-</b>	<b>47</b>
<b>Total (aggregated)</b>	<b>39 (83%)</b>		<b>8 (17%)</b>		

\*) These are again all cases with clear information concerning prices and with different versions.

So, again we can state that commercial publishers do not give up any of their claims but make use of the new medium and market the same way they have been using publishing by print. Above all, there is no visible trend to free access, although in three cases there is a free internet version with an additional print version liable to charge.

With the non-commercial publishers, the tendency towards free internet access is obvious. Under the heading “*No Difference between Internet and Print*”, *free of charge* and *liable to charge* are not much apart from each other (9 vs. 6 cases, with a prevalence of the internet, though), but in all 5 cases with a difference, the less expensive internet version is the free one, not the traditional print version.

Just a last word about journals, I have no quantitative data about print journals, which were completely suspended within the last years and as a possible consequence of the internet. But my colleagues who are in charge with the observation and documentation of journals could not confirm that this was really the case, and if it was it could only hold for single cases, but is definitely not a trend.

So I think as far as journals are concerned, there is no considerable change between free and charged journals. By no means we can say that the border between white and grey literature is shifting or that there is of a mentionable increase of grey journals, neither in print nor in digital form. For both groups - commercial and non-commercial publishers - the internet is the ideal medium for the best possible realization of their purposes:

- staking their claims on a new, growing and attractive market and at the same time lowering costs of production and distribution on the side of commercial publishers,
- distributing their information and ideas to a wider public and doing so much faster, much easier and much cheaper on the side of the non-commercial publishers.

## Chapter 2: Hypotheses concerning grey literature & the internet

For the rest of the publication system I do not have such a comfortable basis as the journals' database. But I have been involved in two research & development (R&D) projects for more than two years now covering all sorts of scientific literature and various parts of the system of scientific information, from commercial publishers via libraries and database producers to research institutes. So I think that our experience is representative for the situation of Social Sciences in Germany, though not quantifiable.

The first of these projects is called *infoconnex*, the second *vascoda*.<sup>3</sup> *infoconnex* is a German R&D project combining three leading information centres for the Social Sciences, Education and Psychology and the three special libraries for these disciplines. The idea of *infoconnex* as an *information network* is rather simple:

- The databases of the information centres can be searched in one single process, i. e. with one single search question although the databases are indexed in different ways, i. e. with different controlled terms, classification codes and the like. *infoconnex* automatically transforms the search terms from one database into the pertinent terms of the two other databases.
- The user makes his choice and can immediately either download a digitalized version of the publications he wants or order a copy be it as a file in pdf or multiple page tiff format or order a hard copy, notably in the case of books.
- *infoconnex* is a *one stop shop*, i. e. the user has only to pay once and can make all his orders in one continuous session. And he can pay either per view or subscribe for a certain time which ranges from 24 hours to one year.

*infoconnex* is just one of four information networks that were built up simultaneously: for *Engineering and Physical Sciences*, for *Life Sciences*, and for the *Humanities*. These four information networks were combined under one common 'roof' called *vascoda*. (Meanwhile, a lot of further institutes, projects and networks have joined *vascoda*.)

In the following part, most information about white and grey literature, about the behaviour of the different 'actors' in the field of scientific information - i. e.: publishers, research institutes, scientific authors, data base producers and so on - are based on (informal) experience in the context of

- the *infoconnex* and *vascoda* projects and
- the Information Centre for the Social Sciences and its different databases concerning literature (SOLIS, ZEITSCHRIFTEN), research projects (FORIS) and research institutes (SOFO).

We can split off our general expectation concerning the future development of white and grey literature and their shares in scientific information into a set of hypotheses.

- **Hypothesis No 1** reads:  
"Digital publication of grey literature via the internet is increasing."

This is undoubtedly true, but the only thing it proves is that the internet is an accepted and attractive medium for dissemination of research findings. And this evidence has nothing to do with structural changes but only with a change of publication media.

- **Hypothesis No 2** is more interesting: It reads:  
"The total number of grey documents ('titles') is rapidly growing, notably of digital grey documents ('titles')."

At least in the Social Sciences, there is no hint that this could be true. The IZ, e. g., has a documentation of grey publication series published by research institutes. Some of them are published in a digital version, meanwhile, in some cases the printed version was definitely given

<sup>3</sup> <http://www.vascoda.de/>; <http://www.infoconnex.de/>

up. But concerning the number of documents ('titles') there was no remarkable, no *countable* change, yet. The documentation of white and grey literature, as well, does not report an unusual increase of grey documents.

- And as a logical consequence - which is **Hypothesis No 3** -: There is also no mentionable change of the quantitative percentage relationship between formally and informally published literature, be it printed or digitalized.

Such a rather firm demarcation between white and grey literature is only partly a question of counting, partly, however, it is a question of attitude or preoccupation. A project like infoconnex, e. g., could be a very good basis for overcoming the obsolete distinction between different types of literature, but no one talks about such a possibility, not even about the idea. Another R&D project called SozioNet (which is to be integrated into infoconnex later on) organizes a cooperative network of research institutes producing digital grey literature and indexing it according to the Dublin Core set of *metadata* (i.e. for instance: title, creator, subject, date, type, format etc.)<sup>4</sup>. This is very interesting since it proves a general agreement of all parties:

- The *project SozioNet* requires the distinction between grey and white literature and its persistency.
- The *research institutes* do not only agree but also invest a lot of work in indexing their grey literature in accordance with the Dublin Core standard.
- The *Dublin Core* itself, which is the result of a *strong international movement*, is evidence for the general attitude that grey literature is something special and the expectation that grey literature is here to stay. By its activities, this movement even consolidates the difference between grey and white.

This leads us directly to the next hypothesis:

- **Hypothesis No 4** reads: 'Personal publication' by the scientists themselves as authors, publishers and distributors are increasing rapidly. And this is combined with a decline of formal as well as qualitative standards in favour of growing wild. So one could say, that grey literature is becoming even more greyish.

Again, this turns out to be not only false but even misleading. Quite in contrary, we can rather state kind of taming of digital grey literature, as we have seen with the Dublin Core initiative which is imposing formal standards to grey literature without, however, denying or only questioning its grey character.

But it is not only metadata, which have been taken over from the formal publication (and documentation) system. A mentionable number of research institutes have already developed internal refereeing systems for papers published in their grey institute series or journals. And it seems, that this is a growing number, though growing slowly.

- **Hypothesis No 5** is my last one: "Everything is free in the internet!"

In fact, this is rather a general assumption about the internet than a hypothesis. Or better: It is a general and widespread error. Maybe this was right in olden times when digital freaks were the predominant species populating the newly discovered continent called internet and for some time defining ideology and style of thinking and behaviour.

But these times are long gone, and only the ideology of total freedom still lingers on. Meanwhile, the internet has become a huge market place with lots and lots of small stalls and big warehouses, as I have already explained above.

<sup>4</sup> see e. g. <http://dublincore.org/documents/dces/> or the homepage <http://dublincore.org/>

infoconnex itself is a typical example for the real situation. Its basic concern is that articles from commercially published journals - be it genuine e-journals or digitalized print journals - are *sold* to users for a price determined by the publishers themselves. So there is no bargaining, neither between infoconnex and the publishers nor between infoconnex and the users. The latter have to pay the price or they do not get admission to the texts. Moreover, infoconnex takes care that existing contracts between publishers and universities are not affected. Nevertheless, the publishers hesitate to sign contracts. Maybe they are still under shock of the slogan of a free internet.

### Chapter 3: Theoretical Explication

If we summarize all the facts I have presented on the last pages, we must say that only little has changed. Above all, the enormous chances for free information and communication provided by the internet are not used within the scientific system. It may be self-evident that commercial publishers have no interest in free publishing since it takes away the basis of their business. But why do scientific institutes not make use of the opportunities presented by the internet? And why, for heaven's sake, do even the scientists refuse to adopt the internet which could be the perfect medium for publishing without restrictions and conditions, for reaching a nearly unlimited worldwide public?

What are the reasons for this inertia of science and the scientists? Is this only conservatism or hostility toward new technologies? Or are there perhaps hidden disadvantages which disqualify the internet as a medium for scientific use?

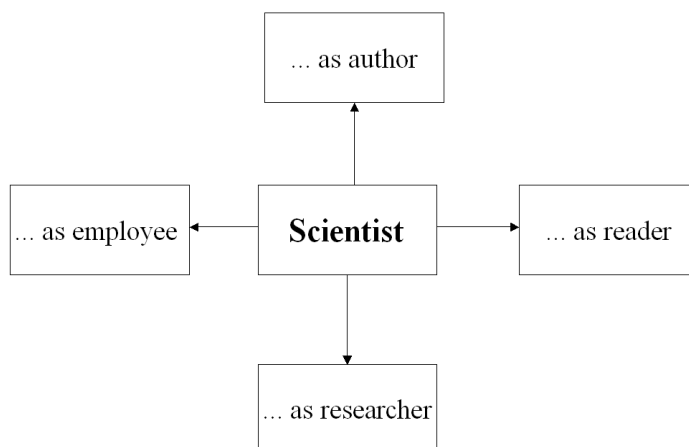
Maybe there is only something wrong with our perception. And maybe we are asking the wrong questions and are making the wrong presuppositions.

### *The Scientist*

Why, e.g., should the scientist really be interested in a barrier free access to publication media? And why should he be interested in barrier free access to all publications of his colleagues worldwide?

Perhaps we can better understand these somewhat heretical if not silly questions if we try to understand the different aspects of a scientist (*see figure 1*):

**Figure 1: Four Sides of One Scientist**



These different aspects are connected with different expectations and requirements. The scientist-as-author, e. g., should be interested in publication media which are barrier free and working fast so that he has his articles and other texts published without delay. And he should be interested in an uncomplicated, barrier free - and that means: free of charge - worldwide access to his publications.

But these are only his supposed interests at first glance, as we shall see in a few moments.

Let us only take the scientist-as-reader or the scientist-as-researcher. As such, he is interested in free access to his colleagues' publications - but only to a certain extent. The total scientific production worldwide is much too much to read, so he needs a mechanism for reduction of this overwhelming complexity of information. But this must not be an arbitrary reduction but should guarantee that the remaining publications are *the most important ones*.

A publication system with free access to all authors, however, could never provide such an opportunity. Everyone can use the internet for his publications, and so there is no pre-selection according to high or low quality, no distinction between important and trivial, and no control whether a text satisfies the standards of a discipline or does not.

A decisive part of such a control of quality and thus systematic and rational reduction of complexity is performed by the formal publication system with its refereeing system. Only a certain percentage of papers is accepted for publication in a special journal (or edition), and publication in this medium is a proof of quality.

But this is not only a reduction of complexity. A severe refereeing process leading to a high rate of refusal raises the *attraction and reputation* of the journal on one side and of the happy few authors printed there on the other side.

So, at the second glance, the interest even of scientists-as-authors is not a free and uncontrolled opportunity of publication, but having their papers published in journals with high reputation (which unfortunately implies a high rate of refusals and a high risk of failure, as we have seen). Only in this case, the scientist-as-author can be sure to acquire scientific reputation himself - and being read by colleagues who are interested in pre-selected and quality-controlled articles.

So, the circle is closed. Where at first glance seemed to be different interests, we now recognize an identity of interests.

At last, there is still the scientist-as-employee. This is an aspect, which is usually completely ignored in Science Studies (or in the Sociology of Science). But in fact the scientist - like any other - depends on his job because he needs to earn his living. So it is important to have and keep his job as a scientist, and in many cases it is also important to receive research funds in order to continue researching, i. e. working, i. e. earning his living.

Here, as well, is a simple relation: If a scientist applies for a job, his future employer looks at his list of publications and puts special emphasis on famous journals (or editions) transferring their reputation to the scientist. And if a research institute applies for a funding of a research project, the same happens on a higher level: What is the reputation of the institute and the researchers? What is the reputation of former research projects? And so on.

Thus, reputation is the universally accepted currency of the scientific system.

### ***The Scientific System***

But I can soothe you: Reputation is not everything. That would mean that only formally published 'white' literature is valuable and that grey literature is nothing.

In fact, the different ways of publication do not depend on the technical devices of an institute or a scientist-as-author, but on different *functions* of document types and publication modes. (And that is why we can be very calmed by the findings presented here. They do not prove that scientists are old fashioned but that they are able and willing to preserve essential scientific standards.)

I have always talked about the scientific system - as if there was only one. In fact, however, there are (at least) four different though highly interdependent ones:

- the *cognitive system* responsible for information and communication (which implies: reducing information to a reasonable and processable dimension);
- the *social system* constituting importance and reputation not only of scientists and journals and institutes but of scientific findings and theories as well;
- the *economic system* comprises commercial publishers and the (commercial) market for scientific information as well as science-as-a-job and scientists-as-employees;
- the '*political*' system which is responsible for research funding and support of research institutes and the like.

Above, I have talked about processes and structures belonging to one of those systems. But as we have already seen in the analysis of the fourfold scientist, there are interrelations between the different systems. For instance, the relationship between cognitive and social system of science is quite evident: Control of access to publication media by a system of referees is obviously a part of the *social* system, and reputation of journals as well as of publications and of scientists is a social factor either, but nevertheless, this has decisive consequences for the *cognitive* system of science because it allows a classification of publications and scientists according to their quality, dignity and importance.

So, publishing in commercial journals with high reputation is important for social and cognitive system alike, but in a certain sense it collides with the requirements of the *political* system, which has to do with research funding. To give an account, it takes completely different ways of publication. At least a part of the papers must be submitted at certain times - in most cases annually -, and must give account of the research project's progress. In certain cases, even the number of recipients and of copies is strictly limited, and a further dissemination requires the permission of the funding institution. So, this special literature requires a special treatment and special ways of control and distribution, which are completely strange to the commercial system of publication. For these reasons, so-called grey literature was invented.

These and the other examples I have given above spotlight only a few relations existing between the highly interdependent parts of the scientific system.

This extremely complex total system of science is the result of a development that has taken centuries. And now, without a rupture it was brought across to the internet; at least it seems so. The scientists (and the commercial publishers), however, had enough creativity to explore and use new and additional opportunities of the new medium for their special purposes, without, however, endangering the whole system of science and its conditions of activity.

The *technical system* called 'the internet' is only a device, nothing more. What really happens depends on the structures and functions of the scientific system and its requirements, not on the technical surrounding.

So there is no reason for blaming scientists or publishers for being old fashioned or hostile against technological progress because they resist the temptations of a free and anarchic internet.

Quite in contrary:

I think, they have been doing a great job.

---

## Author Information

### Dr. Helmut M. Artus

IZ, Informationszentrum Sozialwissenschaften

Abt. Informationsdienste & Marketing (IDM)

Lennéstrasse 30

D-53113 Bonn

Germany

ar@bonn.iz-soz.de  
www.gesis.org/iz

## **Four Winds on the Grey Landscape: A Review of Four Information Professionals Their Work and Impact on the Field of Grey Literature**

Dominic J. Farace and Jerry Frantzen  
GreyNet, Grey Literature Network Service

### **Introduction**

In December 1993, the First International Conference on Grey Literature entitled “Weinberg Report 2000” was convened here in Amsterdam. A special tribute was paid to Alvin M. Weinberg, who in the early 60’s chaired President Kennedy’s Science Advisory Committee and produced the famous report “Science, Government, and Information: the Responsibility of the Technical Community and the Government in the Transfer of Information”.

Along these lines - a decade later - a special tribute is paid at this Fifth International Conference on Grey Literature to four information professionals, who have had long careers in information, who themselves attended and presented papers at the First International Conference on Grey Literature (GL’93), and who have made lasting contributions to specific areas in the field of grey literature. The four are Vilma Alberani, Italy; Peter Auger, United Kingdom; Ulrich Wattenberg, Germany; and Andrei Zemskov, Russia.

Early this summer, each of the four was contacted via GreyNet’s office and asked to provide bibliographic, autobiographic and other human-interest reference material in an effort to gather firsthand information for the paper and its subsequent presentation.

### **Grey Matters in the World of Four Information Professionals**

In the 60s these four individuals were starting their careers in the field of information and science. It was a time when the sky was the limit (virtually space), but when the awareness of unchecked and excessive expansion emerged as smog before ones very eyes. Great challenges offset by (un) known consequences required the skills and foresight of professionals in designing and facilitating information and knowledge transfer between and among the sectors of business, academics, and government. It was in the mid-70s that Peter Auger completed his first publication, which would eventually move on to a 4th edition. The grey primer as it were first entitled *Uses of Reports Literature* revised to *Information Sources in Grey Literature* in its 2nd edition. In the early ‘80s it was with the appointment of Vilma Alberani as Director of the Editorial Services at the Italian National Institute of Health (ISS), where she was able to influence both national and European information policy for the benefit of grey. In the early 90s, Wattenberg returned after two decades of service in Japan turning to the synthesis of his work and that of GMDs initiative to harvest Japanese STI making it available to the West and the rest of the world. This culminated in his edited oeuvre *Japanese Information in Science, Technology and Commerce*. And now, in this, the first years of 21st Century, Andrei Zemskov, stands in the forefront of open source archives, initiating a Russian-German Project to establish a digital archive of Russian mathematical publications entitled *RusDML*.

Each of the past decades, faced with new challenges, witness to scientific and technical advances embraced by these four information professionals.

### **Alberani, the European - Notus, the South Wind**

Vilma Alberani is perhaps the one person most responsible for organizing a national program in and for grey literature in which collection development, processing, diffusion on the one hand were balanced by reflection, authorship, and dialogue on the other. Vilma’s concerns with the implementation of ISO standards for Report and Theses presentations, Report numbers/codes, were among the topics at the National Conferences, which she organized in 1992, 1996, and 1999. Throughout her career, Alberani has (co) authored some eighty articles and fifteen books



and is responsible for the editing of about 250 reports. The turning point in her +40 year career, which started in 1958, came however in 1978 with her invited attendance to the York Seminar. On 13-14 December 1978, thirty participants from European Member States met in York, United Kingdom to discuss ways of improving access to and bibliographic coverage of non-commercial published materials. This meeting was organised by the EEC and hosted by British Library. Among the lasting accomplishments of this seminar were the establishment of EAGLE, the European Association for Grey Literature Exploitation, which celebrates this very month its 25th Anniversary, and its SIGLE-database (System for Information on Grey Literature in Europe), which is today online available via STN International, and which will be covered in one of the P&S Reviews tomorrow.

Vilma returned to Rome after the York Seminar to implement and develop standards for grey literature that would be used not only by ISS, but also by the Italian National Research Council (CNR) and the Italian Library Association (AIB) - in essence in every national center in Italy. No doubt these efforts were responsible for her earning in 1993, lifelong Honourable Membership in the Italian Library Association (AIB). Vilma's work in GL did not end with her official retirement in 2000, but has instead led her to concentrate on information and knowledge transfer to colleagues just starting out and others long working in the field. Within the AIB, she has organised some courses on grey literature and continues to maintain active e-dialog.

### **Auger, the Encyclopedist - Zephyrus, the West Wind**

Peter Auger started out with Lucas Industries in 1958 moving up through various departments, while envisioning a mosaic of grey literature collections and services well beyond Lucas. In fact his published work in 1975 and its later revised editions provided the first Roadmap of Grey Literature Systems and Services. His words of preface echo throughout each new and revised edition: "In literature, color coding in the literal sense is nothing new" - white papers (official documents), green papers (proposed government policy), blue books (official reports), black books (e.g. from the national treasury, red data books (lists of endangered species) all of which are grey - i.e. not controlled by commercial publishing. The first chapters of each edition spelled out the ABC&Ds of grey Acquisition, Bibliographic control, Collection development & Distribution. This is then followed by a roadmap for the various branches of science and information, namely Aerospace (e.g. ESA, NASA); Life sciences (e.g. Caricom, Agris, Biosis); Business and Economics (e.g. IMF, WB, UNCTAD); Education (e.g. Eudised, Unesco, ERIC); Energy (e.g. DOE, ERA, IAEA, INIS); Science and Technology (e.g. CORDIS, DTIC, NTIS, VNTIC), all familiar organizations and agencies to this international GL conference series. In 1988, Peter left Lucas to establish his own information consultancy, and has since retired in 1997.

### **Wattenberg, the Ethnologist - Eurus, the East Wind**

After receiving his Doctors degree in natural science in 1972, Ulrich was granted by the German Academic Exchange Service (DAAD) a four-year scholarship to Tokyo University, where he studied Japanese and did research on silicon. It was then that he became aware of the richness of Japanese Scientific and Technical Information. As it might appear, Ulrich was being primed in a field of specialization for which an infrastructure was to be custom built. The name of the organization would change throughout Ulrich's further quarter of a century stay in Japan, where he worked in the liaison office of IDW (Institut fuer Dokumentationswesen) a department of the Max Planck Society, later to become GID, then GMD, Gesellschaft fuer Mathematik und Datenverarbeitung), today known as GMD First. While names changed, the human resource, Ulrich Wattenberg would remain the constant, until his return to the home office in Berlin in 1993. In that year, JICST now JST held membership on the GL'93 Program Committee and they were instrumental in Ulrich's response to the call-for-papers. During the final decade in his information career, Ulrich was confronted with a more formidable change than that of a name. The earlier investment principle applied to grey literature, where there would be "first costs then benefits" was replaced by other principles to which grey literature would have to adjust. Industry was calling for exclusivity, value added and tailor made information. Upon his retirement last year, Ulrich was not as yet convinced that internet search engines, while

providing valuable grey literature would be able to completely replace the contacts and links of the information professional - for him a prerequisite to GL acquisitions and diffusion.

### **Zemskov, the Envoy - Notus, the North Wind**

In 1965, Andrei Zemskov received his PhD in Low Temperature Plasma Physics from the Moscow Institute of Technical Physics. In that same year, he began as scientist, moved to senior scientist later to become Supervisor of the Plasma Physics and High Energy Accelerator Program at the Kurchatov Atomic Energy Institute in Moscow. Andrei was also politically active in the CPSU, first as secretary of a local, later to First Secretary of the Moscow Regional Office. In 1990, just months after the fall of the Berlin Wall, Andrei Zemskov was appointed Director of the Russian National Public Library for Science and Technology a position that he still holds today. His work at NPLS&T and his professional and personal interest led him to explore the free access of information, grey literature, and compact disc technology.

Andrei Zemskov is in more than a way a bridge from the soviet closed society to a democratic open one, where information is free to access, where there are no checkpoints and walled borders on the electronic highway. Andrei's affinity with grey may have been his break with red? He is author of more than 90 publications in physics and LIS (Library and Information Studies) a number of these, which deal with grey literature and which carry the term explicitly in the title. He served on the editorial advisory Board for IJGL and currently is Chief Editor of SciTech Libraries Journal. He has membership on a number national and international Committees such as IFLA, IATUL, ILIAC, and the former FID, etc, No doubt, today is not the first time that Andrei Zemskov has been honored for his achievements. He is the recipient of national and international recognition for Merit, Leadership, and Friendship, for these the Who's Who in the World, 16th ed. will attest. However, now on the eve of his 65th Birthday next month, we pay tribute to his contribution to the field of grey literature. A field, which has attracted and drawn Zemskov, Wattenberg, Auger, and Alberani together; and, which first brought them together in Amsterdam in 1993 and again today, by way of tribute.

### **Close**

What these four information professionals hold in common is there years of tedious and meticulous work in the field of grey literature, above and beyond their organizations mission. Combined they represent the various sectors of government, academic, business and industry i.e. the entire grey spectrum. Thank you Vilma, Peter, Ulrich, and Andrei for the combined 100+ years of work in the information profession serving as ambassadors of grey.

---

### **Author Information**

#### **Dr. Dominic John Farace**

TextRelease/GreyNet,  
Grey Literature Network Service  
Beysterveld 251  
1083 KE Amsterdam  
The Netherlands

info@greynet.org  
info@textrelease.com  
www.textrelease.com  
www.greynet.org

## Capturing Academic Grey Literature - Starting at Home

Gretta E. Siegel,  
Portland State University, Portland, OR USA

***Abstract:** Historically, many academic libraries have shunned grey literature for a variety of reasons. When attention has been paid, academic concerns about collection development and grey literature tend to focus on external collections from industries or agencies and, more often than not, these collections are oriented to a particular subject or discipline. At GL'99, one of the discussion groups talked about making efforts to investigate, collect, and provide access to the scholarly grey output of the universities themselves. A study was undertaken to determine the sources and abundance of scholarly grey literature produced at a single university campus (PSU), and to determine the extent of bibliographic control by the university library. The study involved surveys across campus as well as bibliometric analyses of catalog records including classification analysis and analysis of authoring bodies. The findings of the study were reported to the library staff and administration, resulting in a slight paradigm shift in how the library perceived the academic units of the university, and also in some concrete changes to the library's collection development policy and procedures.*

### Background and Introduction

In academic libraries, discussion about creating bibliographic access to grey literature conjures up responses ranging anywhere from blank stares, to knowing grins, to terror. Why? The blank stares come from those who have no idea what you are talking about, mostly because it has, in the past, been in the realm of obscure library jargon; the knowing grins come from those who have wrestled with this literature, know its value, but also know the challenges of working with it; and the looks of terror come from those who know what it is in its broadest sense, and equate it with Pandora's box. One reason, even within the world of libraries, for these varied responses, is that this is not a subject generally dealt with in formal library training. Excellent cases have been made for the inclusion of this area in an LIS curriculum, but we have not seen its explicit integration to date.<sup>1, 2</sup>

Another part of the problem is definitional. The widely accepted definition for grey literature is **"that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers"**<sup>\*</sup>. Imagine - this includes things like instruction manuals for your toaster, bus schedules, tax forms, etc. Thus, the looks of terror. So, when we approach doing anything practical with this body of literature, we need to look at subsets of it, that which interests us, and define it well. In a paper by Gokhale<sup>3</sup>, all of grey literature is considered and a multidimensional detailed analysis of what constitutes 'grey' in various subject areas is provided. The author approaches the definitional problems by presenting a taxonomy of grey literature within different subject areas, and highlights the potential for systematic research in this area. At the other end of the spectrum, a case study presented by Cianelli and Leyton<sup>4</sup> discusses the formation of a regional cooperative information center with a very focused topical scope (Latin American Nursing). They outline concepts to consider for selection criteria as well as policies and standards for collection and management. In a situation like this,

---

<sup>\*</sup> Definition adopted at the 3<sup>rd</sup> International Conference on Grey Literature, Luxembourg, November 1997.

the definitions come much more easily as the stakeholders, the users, and the scope of the collection are fairly obvious. While these articles and others approach the topic from a subject orientation, there are occasions when other criteria, independent of subject, may play a stronger role. In getting 'buy-in' from academic colleagues to embark on a grey literature project - clear articulation of definitions and scope are essential.

In university libraries, many categories of subject independent grey literature are collected; they just aren't thought of as being 'grey'. For example, student theses and dissertations are collected, as well as quite a lot of government documents. Scholarly conference proceedings in subject areas of interest are also collected, and when looking at the academic environment in aggregate, it would not be surprising to find that better than 95% of these are held in college and university libraries. Many professional society publications are also technically considered 'grey' under the present definition. Libraries, however, would rarely make this distinction. These are just several examples of things that most academic libraries think of as just the normal literature that needs collecting, and it does not tend to be seen as 'special'.

Other types of scholarly grey literature, i.e. that which is produced as a result of a scholarly study or inquiry, but which is not published through the traditional channels of books or journals (commercial or society), are also produced on academic campuses and encompass much important work. Since it receives limited distribution, it is often not given an opportunity for wider dissemination through traditional bibliographic treatment, i.e. collection, cataloging, and inter-library lending. However, this material does on occasion get cited in scholarly works. This, in turn, leads subsequent researchers down the very frustrating and often futile path of location and procurement. At the GL '99 conference in Washington D.C., there was some informal discussion about what academic librarians could do in an effort towards closing this bibliographic gap. The study reported on here is an effort in that direction, and is a case study of one university. The methods presented here could easily be applied to other institutions.

The institution is Portland State University (PSU) in Portland, Oregon. It is a doctoral granting, comprehensive university with approximately 15,000 FTE students. The operative definition for the study was: that which is produced **BY** faculty or staff **IN** the university, **FOR THE PURPOSE** of sharing scholarly information with others. This definition precluded the consideration of many things produced in academia which would be more appropriate to a university archive, or which would be seen as ephemeral, or which would open up the infinite realm of student produced literature. The study encompasses two different assessments, one is an investigation of the scholarly grey literature produced on the PSU campus, and the second is an assessment of how well we are presently providing bibliographic access to this body of literature.

### **Methodology for Assessing Production of Grey Literature**

A survey was prepared and distributed to all academic departments, programs, centers, and institutes on campus. The survey is shown in Figure 1. The introductory paragraph of the survey clearly defines the type and source of material that the study is

**Figure 1: Survey Instrument**

March, 2000

Library Survey for Scholarly Grey Literature

We at the PSU library are interested in publications produced by your department, program, school, center, or institute. We are seeking scholarly or technical reports produced by regular faculty or staff, which are published here at PSU and intended for limited distribution. This would include conference papers that have been published in full proceedings of meetings, but which the library may not have acquired.

If time and funds permit, we would like to collect this material and add it to the library collection so that it will be available to students and researchers. Please note that we are NOT interested in materials of an ephemeral nature (e.g. brochures, newsletters, administrative notes or memos, etc.), or in materials written by students or interns.

We would appreciate it if you could take a few minutes to fill out this questionnaire. Please see the reverse side for examples of appropriate items. Thank you for your assistance.

1. Name, title, and e-mail address of person completing the survey:
  
2. What is the name of your department, school, program, center, or institute?
  
3. Do you produce any reports of the type described?    Yes    No  
 If so – could you please give us the titles and authors of individual reports, or, the title of the series and an estimate of how many separate items there are within the series?  
 (use a separate page if necessary)
  
4. Do these exist in paper format, electronic, or both?    paper    electronic    both
  
5. For the ones that exist in paper, would you be willing to donate 1 copy of each report to the library?  
**Yes   No**
  
6. For the ones that are electronic, would you be interested in working out an arrangement with the library to create access to them?    **Yes   No**
  
7. Please list a contact person willing to coordinate obtaining these reports from your department:
  
8. Any comments you would like to share with us?

Thank you very much for taking your time to help us with this project. Please return to your library liaison or to Gretta Siegel, 5<sup>th</sup> floor Millar Library, mail code LIB-W, by March 24, 2000.

gs 3/00

seeking to uncover. The reverse side of the survey gave examples of appropriate reports to be included.

The list of entities to which the survey was sent was derived from the current campus directory. Entities for which there was adequate doubt that they would have the type of materials desired were not surveyed, for example, the athletics department or the registrar's office. On the other hand, the university's many programs, centers and institutes that are separate from academic departments, were included. Sixty surveys were distributed. Distribution was done with the assistance of all of the public services librarians since it was thought that their liaison relationship with the departments would result in a higher response rate.

Returned surveys were tabulated to show how many documents were reported by each entity. For those that provided lists of documents, the items were checked against the library catalog to see if the library already held them or not.

### **Methodology for Assessing Bibliographic Control of Grey Literature**

The library's online catalog was examined for records that met the same criteria as that which was defined in the survey, i.e. scholarly in nature, produced by faculty or staff, not ephemeral, and intended for an audience beyond the walls of the University. Several steps were necessary to derive a list of materials to be counted and analyzed. The librarian in charge of catalog maintenance performed searches on the 'back-end' of the catalog. Reports were created based on searches of the author, corporate author, publisher, and imprint fields. Several searches had to be run and reports created in order to catch as much as possible and at the same time to eliminate theses and dissertations, since these are produced by students. The resulting reports were then manually reviewed and edited for which items to include or exclude. Examples of items which were excluded are: class schedules, yearbooks, newsletters, syllabi, workshop outlines, reports written by student interns, budget reports, university self-studies, enrollment reports, etc.

The remaining items were then analyzed with regard to subject (via Library of Congress classification numbers) and with regard to what campus unit produced them.

### **Results**

Twenty-three surveys, or approximately 38%, were returned. Since overlap between some departments, centers, programs, and institutes exist, and because some responses came from units who were not originally identified, it was expected that there would not be an exact 1:1 correlation between surveys distributed and those returned. When analyzing the holdings of the library catalog, a few units who had not been surveyed were found to have produced some material that met the criteria of the study, thus also contributing to the lack of a 1:1 correspondence. Due to these factors, quantification of results was approximate, but adequate for the observation of trends and patterns.

From the information returned on the surveys, a sample (60 items) was searched on the library's catalog to get an idea of how many of these documents the library already owned. It was found that the library owned 11 of these items, or 18%. Some of the reports listed were found to be duplicates. This is due again to the fact that there is some overlap between some of the units. The degree of duplication was estimated at ~10%. The overall estimate of the percentage of reports produced on campus that are cataloged into the library collection is 16.5%. This figure was derived as follows:

$$.90 \times (\text{unique items in list}) / \text{total items reported} \times 11 \text{ items cataloged} / 60 \text{ unique items in list} = 16.5\% \text{ cataloged items} / \text{total items reported}$$

Most of the items that were reported on the surveys that were already owned by the library were older reports. Another observation was that the Computer Science and Civil Engineering departments produced over half of the items reported on the surveys. Other than these, however, most of the grey literature reported to us was being produced by centers and institutes, rather than by academic departments.

The results of the catalog analysis of items already held by the library yielded 242 unique documents. Analysis by originating unit showed that the centers and institutes were the primary originators of grey literature on campus with 68% of the reports

coming from them, as compared with 9% coming from academic departments or programs and 7% coming from service units, such as academic computing or the library.

An additional analysis was done to find out if there were any subject patterns for where this collection of literature falls. It was clear that most of the previously collected materials fell into the social sciences, including economics and education (61%), and that very little fell into the humanities (4%) or the sciences, including engineering (10%). A small amount though, did indeed fall into practically every subject area.

### Discussion of Results

The 38% return rate on the original surveys was lower than desired, but high enough to yield very useful information and was enough to note the pattern of where this literature is being produced on campus. Even though this return rate would be considered reasonable for an average survey, the goal of capturing all of this literature on any campus would necessitate a 100% return.

The results of the 'originating unit' analysis, showing that the great majority of these documents are produced by 'Centers' and 'Institutes' on campus, rather than by traditional departments, brings to bear an important issue regarding collection development and issues of library liaison relationships. In this particular library, as in many, librarians have liaison duties with academic departments. These relationships are invaluable for collaboration over purchasing decisions, as budgets are often allocated on a departmental model. Since, in the case of this institution, acquisitions budgets had not historically been allocated to centers and institutes, there had never been formal liaison relationships established with these entities. Since this is where most of the grey literature is produced, it should be no surprise that the acquisition of these materials has been hit or miss, and that the percentage of the available documents that are cataloged is quite low.

At this point, the question arose as to whether or not the library values, or should value the collection and cataloging of this material. The evidence that a sizeable amount of it was found in the catalog (mostly older) demonstrated that at one time, these materials were worth capturing. A discussion with the library faculty on our campus resulted in consensus that they still are.

### Collection Development Implications

As mentioned in the introduction, any approach to the study and management of grey literature must take definitions into consideration. Two good articles that go into considerable detail in defining and explaining grey literature are Schmidmaler<sup>5</sup> and McDermott<sup>6</sup>. While McDermott tends to cast this subject in the light of a problem to be solved, Schmidmaler paints it more as an opportunity to engage with a broader spectrum of literature. While I prefer the more positive outlook, one cannot disregard that those who are charged with the tasks of acquisition and cataloging will tend to see it more as a problem. This is where it becomes helpful to develop clear collection development policies, as well as guidelines for acceptable cataloging that are perhaps less rigorous than what might be applied toward more mainstream or 'white' literature.

In an article by Alison<sup>7</sup>, the argument is made that vendors could organize bodies of grey literature in a way that would be marketable to libraries, and then libraries would buy it because it would be on the acquisition 'radar screen'. While this idea has merit, it presumes that the materials we want to collect are all generated externally, and that all grey literature has 'marketability'. In fact, all of the literature on collection development of grey literature seen so far, discusses bodies of literature produced external to the academy that need to be included in subject collections.

While internally produced academic grey literature does not likely share the marketability of certain agency produced scientific reports (the sort that would lend itself possibly to Alison's proposal), it is still valuable scholarly output that needs to be collected. Therefore, other ways are needed to get it onto the 'acquisition radar screen'.

Most grey literature collection development discussions, besides presuming interest in only external documents, center around the pros and cons of creating special collections vs. integrating grey literature into the larger library collection<sup>8,9</sup>. Gelfand<sup>8</sup> notes the trend away from forming special collections, as well as the absence of any grey literature collection development policies in the university system where she works (University of California). For many reasons, including collection management and usability, I am a strong advocate for integration. All grey literature, and in particular, the type uncovered by this study, does not fall into tidy packages that would lend themselves to special collection treatment. If material is worth collecting and worth cataloging, then it should be as easily located as anything else held by the library.

As a result of this study and of advocacy on this issue, our library has made some positive changes. We have now enhanced the roles of the librarian selectors by assigning liaison relationships to each Center and Institute on our campus. It will be part of one's collection development duties to maintain an awareness of any reports produced by these units and to collect them (whether in print or electronic) and get them into the cataloging pipeline. To avoid uneven collecting practices, we have added a page to our collection development policy manual that outlines criteria for grey literature selection and acquisition. The additional policy statements are shown in Figure 2.

### **Concluding Remarks and Areas for Further Investigation**

My interest in capturing grey literature dates back to the mid-1980's, though it has never seemed to generate much excitement among most of my professional colleagues. The advent of the web, however, created the ultimate in grey literature - millions of ephemeral websites. The late 1990's saw several massive, and laudable, efforts launched at cataloging the web, both the visible and the invisible. This struck me as ironic, especially because the people who wanted to embark on this ambitious task, were often the same ones who did not see any point in dealing with paper based grey literature. While I learned to keep this observed contradiction more or less to myself, I was happy to see this same observation ("Isn't the Web just a huge pile of gray literature?") made in a recent article.<sup>10</sup> Thus, whether one is dealing with the Web or with more traditional formats, it gets us back to the necessary step of articulating definition and scope of what it is that we want in our collections.

All grey literature is not worth collecting, but much that is worth collecting, is still falling through the very large cracks of standard library acquisition practice. The ease of 'acquiring' access to web documents easily outstrips that of acquiring other grey literature, and this may, in part, account for the enthusiasm towards the practice of cataloging the web. This acquisition dilemma highlights yet another type of 'digital divide' that is worthy of some attention.

Another consideration is that university libraries are increasingly asked to collect, catalog, and house grey literature collections that are externally produced, though of interest to the primary and secondary clientele of the library. While in many cases, this is an excellent service to provide, it seems that the first step in such a program should be to capture that which is produced by the home institution. Since it cannot be anticipated that some 'other' university will be interested in collecting all that is produced on one's campus, it is important that university libraries capture as much of



this locally produced scholarly literature as possible, regardless of format. If this is done, inter-library loan departments everywhere will be grateful. Once this kind of housekeeping is in order, the library is then in position to consider community-based grey literature collections that may be appropriate to the mission of the university

A third consideration is of course the collection of suitable scholarly output generated by students. In this realm, well-crafted definitions will be of extreme importance. Depending on the size and scope of these collections, it may be most suitable for this body of literature to be captured electronically through the increasingly popular 'web portals' that are appearing on university websites.

### Figure 2: Collection Development Policy Statement

V. Institutional Scholarly Grey Literature: It is within the mission of the library to capture, preserve, and make available the scholarly output of the institution. To this end, the library will attempt to acquire technical reports and other scholarly publications produced by PSU Departments, Programs, Centers, and Institutes. These materials will be cataloged and added to the collection, whether in print, electronic, or both. Criteria for selection is as follows:

Authorship: The primary author(s) should be PSU faculty or staff

Content: The content should be such that a person doing scholarly research might choose to cite the work

Publication: The item would generally not be published commercially, but would have been produced in a quantity intended for limited external distribution.

Examples: Technical reports, reports of studies, conference papers that have been published in full proceedings of meetings, but which the library may not have acquired.

Examples of what NOT to collect: Materials of an ephemeral nature (e.g. brochures, newsletters, administrative notes or memos, workshop notes, course schedules, etc.); materials written by students or interns.

## References

1. Julia Gelfand, "Teaching and exposing grey literature: what the information profession needs to know: examples from the sciences." *Third International Conference on Grey Literature, November 13-14, 1997, Luxembourg*, (Amsterdam: GreyNet, 1998), pp.242-50.
2. L.O. Aina, "The management of grey literature as a component of a library and information science curriculum." *Third International Conference on Grey Literature, November 13-14, 1997, Luxembourg*, p.251-8, (Amsterdam: GreyNet, 1998), pp.251-8.
3. Pratibha A. Gokhale, "Grey literature varieties - definitional problems." *Third International Conference on Grey Literature, November 13-14, 1997, Luxembourg*, (Amsterdam: GreyNet, 1998), pp. 259-73.
4. Rosina Cianelli and Virginia Leyton. "Criteria for selecting grey literature." *Proceedings of the Second International Conference on Grey Literature, 2-3 Nov. 1995, Washington DC*, (Amsterdam: GreyNet, 1996), pp. 233-7.
5. Dieter Schmidmaler, "Ask no questions and you'll be told no lies: or how we can remove people's fear of 'grey literature'." *Libri* 36 (1986), 96-112.
6. Judy C. McDermott, "Defining grey literature for the 21<sup>st</sup> century." *Proceedings of the Second International Conference on Grey Literature, 2-3 Nov. 1995, Washington DC*, (Amsterdam: GreyNet, 1996) pp.9-17.
7. Peter Alison, "Stalking the elusive grey literature." *College & Research Libraries News* 48 (1987), 244-46.
8. Julia M.Gelfand, "Academic libraries and collection development implications for grey literature." *Publishing Research Quarterly* 13 (1997), 15-23.
9. Geoffrey E. Cleave, "Project management in grey literature; a study based on the development of the Working Papers Project at the University of Warwick." *Second International Conference on Grey Literature, 2-3 Nov. 1995, Washington D.C.* (Amsterdam: GreyNet, 1996) pp.74-82.
10. Andrew K. Pace, "Black, white and shades of gray (literature) on the web." *Computers in Libraries* 22 (2002), 44-47.

---

## Author Information

**Prof. Gretta E. Siegel**  
Branford P. Millar Library  
Portland State University  
PO Box 1151  
Portland, OR 97207-1151  
USA

siegelg@pdx.edu  
[www.lib.pdx.edu/about/faculty/siegel.html](http://www.lib.pdx.edu/about/faculty/siegel.html)

## The Ubiquity of Grey Literature in a Connected Content Context

By Julia Gelfand  
University of California, Irvine Libraries

### ABSTRACT

Ubiquitous or pervasive computing has become a new paradigm in computer science and networking and seeks to embed computers into all functions of daily life, where the technology may be invisible. This "calm technology" allows for some powerful new products and ways of doing things. In the library, publishing, and information arenas this may mean potentially transforming the bibliographic organization and how scholars find and use information. The social and cultural enhancements ubiquitous computing offers is very pronounced in information rich sectors. Grey literature is an example of nontraditional publishing and creative output that has benefited from the attention of how new models in human-computer interactions and computer supported cooperative work blend many different disciplines, contribute to new emerging areas and record that evolution. This paper will explore how context-aware computing supports specific products such as grey literature. Applying the work of Mark Weiser, Paul Dourish and Anne Galloway and reviewing hybrid applications that incorporate new interfaces, displays, operating systems and wireless communications we can explore where and how grey literature fits in and what may soon become possible for a new generation of grey literature.

### Definitions and Relationships

This paper was to have been the last paper delivered in the last session of a conference celebrating a decade of work in grey literature and you ask why all this emphasis on ubiquity as a frame of reference for new models. So it may not be a perfect fit in this session which is not about roadmaps for the future, but we will take a journey anyway and try and make it fit in this session about models of grey literature. Well, in this last decade we have moved from celebrating unique elements in grey literature to seeing and concluding that more is grey than less. We also have observed and experienced that the availability of information and instructional technologies and the changing landscape of organizational infrastructures and cultures have directed and influenced us and our colleagues, especially in the sciences, to researching, composing, and disseminating our work in very individualized or institutionally branded packages rather than in bulk publishing outputs, once the milestone of journals, books and monographic series. Publishing has changed rather radically and the economics of traditional commercial and academic publishing is challenged because the financial models can no longer be sustained and intellectual property and digital management rights now encourage authors and contributors to retain their rights. We all know that value-added enhancements made by professional publishing enterprises add great benefits, but the digital age demands different controls. The methods associated with online publishing, archiving and discovering reading habits all related to grey literature is what the focus of this paper is about in a connected content context.

### Ubiquitous Computing

Ubiquity or ubiquitous claims to mean "existing everywhere or inescapable" and when applied to computing and technology suggests that the horizon for access and creativity is completely open. Some writers have noted it is descriptive of a paradigm shift "where technology becomes virtually invisible in our lives."<sup>1</sup> There are many different categories of indicators that reinforce the concept of connectivity and networking and can be summarized by access, learning, society, economy and policy.

Since 2000 when John Seeley Brown and Paul Duguid authored the *Social Organization of Information* and the journal, *Ubiquity* was first issued in 2001 by the Association of Computing Machinery (ACM) many of us have been inspired to see how this new niche segment or specialization influences how we use technology and what changes there are in our base of work and operation. There has been a new demand for a sense of community or belonging and affiliation has taken on different parameters. The role of the independent consultant or scientist is more common. Often associated with developments in mobile and wireless computing, ubiquitous computing includes extending physical limits of computing, incorporating multimedia, responding to challenges in privacy and security and treating software engineering issues. Obviously the spectrum will continue to expand with many different applications now birthing themselves.

In a recent book by John Barendse, *Developing More Curious Minds*, I made the analogy that not only do we need more people "who are awake to the possibilities of alternative futures,"<sup>2</sup> but also we need scholars and scientists to be more awake to the possibilities of alternative publishing models and ways to share information. That has always been characteristic of what has defined grey literature. There have been obstacles in how to find and retrieve the literature or content, how it should be organized and controlled in institutional settings such as libraries and how it should be preserved and archived as it goes through revision and change. Grey literature is about creating models of inquiry and schools of inquiry that promote content building, working deeper and perhaps broader but within a sphere of relationships. Thus the title of this paper "the ubiquity of grey literature in a connected content context" applies.

### Searching, Evaluating and Utilizing Grey Literature

By using the Internet or World Wide Web to foster inquiry, investigation, and interactivity and build these relationships, new areas of discovery and curricular interest are available and new reading habits are born. When we teach young students the value of the net we are trying to achieve a sense of independence and sense whether they can evaluate what they find - have they found something important or relevant and how do they apply criteria to that. Traditionally, we teach the elements of audience, authority, purpose, objectivity, currency, coverage, accuracy and relevancy as appropriate criteria.

In a more connected, networked environment, we may extend those criteria to include usability, aesthetic value, comprehensiveness, and connectivity. We want our users to be able to link to opposing points of view and see the spokes of a major issue that may include aesthetic, historical, mathematical, health, moral/religious, scientific, social, economic, philosophical, physical, mechanical, cultural, geographical and personal aspects.<sup>3</sup>

When we think of examples of grey literature and include among them field notes from a geologist, a daily journal from a writer, migrations from an ornithologist, logs of changes from a laboratory scientist, simulations or visualizations from an animator, genetic code for an ill patient, we are working in a new environment other than just descriptive analysis. We are promoting critical thinking and applying the pervasive computing environments that encourage multimedia to synthesize and use by a wider audience than just the creator.

### Contextualizing New Products

Many of us who have studied grey literature as a product of science communication will welcome another new book on communicating science. Scott Montgomery, in the *Chicago Guide to Communicating Science*, does not use the concept grey literature, but devotes a chapter to "other types of writing," where he suggests that review articles, book reviews, debate/critique all have a place in scientific scholarship. As a science journalist, he also attributes how the proposal, graphics, oral presentations and conferences stand-alone and together to contribute to scientific knowledge. In the online environment, he compiles a list of resources or what we define as examples of grey literature and demonstrates their value to understanding and probing science:

- Online journals, newsletters, magazines, publishers
- Preprint archives
- Major bibliographic resources
- Professional society and association sites
- University department sites
- Library sites
- Data archives
- Personal web pages
- Government agency and program sites
- Research program sites
- Industry sites
- Research institutes
- Local scientific society sites
- Observatories
- Image catalogues & archives<sup>4</sup>

Innovation and technology appears to be a hand and glove combination. This means that it is increasingly difficult to think that business or academe can truly practice innovation without a technology process to guide it or keep it kindled or warm. Many business writers and CEOs in recent years have explored how to marry the knowledge industries with innovative business methods and successful products. Knowledge management was born from realizing how important the relationship information has to new product lines and what a large business sector it has become. Schools of business and management around the world focus on what kind of forces are used to generate technology and new ideas and thus new products and services. Investment banking and portfolio contributions have helped to develop viable innovations to take to the marketplace. Richard Leifer at the Rensselaer Polytechnic Institute's Lally School of Management has been exploring how disruptive radical innovation is and how companies can achieve breakthrough innovation.<sup>5</sup> When one considers the power and utility of the world wide web to build, create and demonstrate new ideas and then to communicate that to the world-at-large, that was certainly an example of "thinking out of the box" in the early 1990s. Increasingly, additional complications and complexities figured into the equation such as how to archive and preserve such outputs and new thinking was shared by many very clever minds. Among those, is Esther Dyson who responds to an interviewer in early 2001, "the Internet needs a lot more social infrastructure. Not necessarily regulations, but affordances, directories, reputation systems, mapping systems, better search tools, better ideas for avoiding spam without abridging senders' rights to freedom of speech and so forth."<sup>6</sup>

If creating new business models are as important an innovation as products themselves, then grey literature has a very strong future. Communications and now wireless technologies afford both creators of new products and users or subscribers ways to organize information. Thus scholars and scientists have access to a wider range of resources and a new sense of competitive intelligence emerges among those users. Internet technologies are responsible for how knowledge is created. One of the most cogent and convincing books on this topic is by Francis Cairncross, *The Company of the Future: How the Communications Revolution is Changing Management*. It again, does not refer to grey literature directly by name, but skirts it by stating, "Knowledge becomes accessible in new ways; it becomes easier to store and to transfer; expertise becomes easier to locate; employees can collaborate more effectively, whether they work in the same firm, or time zone or in different ones...technologies must reinforce existing human patterns of knowledge management; they cannot replace them. Technology is only half the answer; managerial ingenuity must do the rest."<sup>7</sup>

### Connectivity and Networking with New Innovation

What this means for grey literature is that computing, communications and connectivity must blend together. If innovation does indeed feed on shared knowledge, then web-based

publishing is what has and will take grey literature forward. Colleagues working simultaneously from different points around the globe can achieve collaboration and development of a product can be modular rather than sequential. The Internet also allows for easy monitoring by competition as well as self-monitoring and tracking of progress that aids the competitive intelligence that innovation must aggressively participate in. Examples of this are rushing to register new patents, protect ideas and utilize open source software. The latter leaves a greater sense of “innovation etiquette,” so that the competition knows all the tricks that have been tried. Having the “right” team of personnel attached to new products also contributes to the likely success with innovation. As Cairncross states, “companies must find ways to ensure that knowledge and ideas bubble up from the bottom, as the clearest symptoms that a company has a healthily creative culture. They need to be constantly aware that the most vibrant and useful ways in which people share knowledge online will often be electronic versions of the ways human beings have always contacted each other and picked each other’s brains.”<sup>8</sup>

Why ubiquitous computing is meaningful for grey literature has its origins in the early work of the late Mark Weiser at Xerox PARC, which was devoted to the social rather than the technological imperatives of technology. Anne Galloway explores this work in a fascinating paper, “Resonances and Everyday Life: Ubiquitous Computing and the City,” and the applications to grey literature are compelling. For starters, she introduces John Seeley Brown and Mark Weiser’s ideas that “calm technology could be distinguished as technology that would be so embedded, so pervasive, that it could be taken for granted. It would be informative without being overwhelming or distracting. They suggested that ubiquitous computing would become so commonplace, so unremarkable that we would forget its enormous impact, just as we have with writing and electricity, two other ubiquitous technologies.”<sup>9</sup> The mechanics and business arrangement of how ubiquitous technologies can be commercialized and sustained is what makes grey literature a very interesting case study.

### Transformation of Content: Challenges of Sustainability

Content that has a beginning, a transition and perhaps a future edition not yet defined and that which will contain links to lateral as well as related, deeper, or even shallower references, a mapping of relationships occurs. To some this may resemble a form of indexing, or establishment of priorities and directions. In the library context, we engage users in a reference interview to learn what the purpose of their research is - for what reasons do they seek the information and how do they expect to use it. By offering directions to primary, secondary, or tertiary sources of information, an information specialist and consumer of information can ascertain the relevance of the match.

### Back to Grey Literature

When one creates content that is expected to be defined as grey, it is assumed that it will lack the mapping that was just aforementioned. In reality, that is not usually the case today, because of the success of search engines and web-crawling devices that use different algorithms and means to discover and retrieve websites created by independent publishers, authors and contributors. Additional ease in finding relevant grey literature is increasingly apparent but the processing and organization and compliance with bibliographic control standards still remains deficient. Thus, the individual responsibility of managing one's grey literature references or holdings is challenging.

Grey literature is the essence of portability. The proliferation and ubiquity of its visibility in all domains of electronic publishing today from all countries of the globe, suggest some common attributes that reside in the functionality of computing. The increasing dependence on more powerful and robust laptops and hand-held devices rather than only on the desktop demonstrates the intense use in instructional and educational technologies. Greater advocacy for educational software and learning applications reaffirm the role of grey content in classrooms from early childhood to the most specialized and sophisticated applications. Digital

libraries have sprung up in educational and commercial settings and distance education was pioneered using earlier forms of grey literature as primary content. Another measurement or benchmark is the increasing citations to content in traditional literature that cites and references the nontraditional or grey literature. Conducting research in most fields today must reflect a literature search composed of all fields. This is predominantly the case in scientific publishing, once the genesis of grey literature and today the mainstream of the movement.

### Current Realities

Earlier this year, a new model releasing scientific information was launched. This is the Public Library of Science (PLOS), founded on the premise that scientific research should be freely available. Defined as open access, the business model is nearly the reverse of what has been done in the past. An "author pays" policy was introduced where the author is charged expenses to cover the cost of the peer review process and production associated with creating an online archive. Different and more expensive to submit to than the British product line of BioMedCentral journals, this new effort encourages an embargo be lifted after six months of released content and deposit it into a public-accessible database which could then be used for text mining. The overwhelming interest in this model was illustrated by 30,000 signatures that encouraged developing this model that made it clear that "author power" can influence new products.

Universities around the world have opened up clearinghouses, repositories, eScholarship collections to help authors organize their work and get recognized within the academic principles of scholarly communication where the tenets of peer-review and high-end scholarship are acknowledged and valued. This would not have taken place without the power of ubiquitous computing and the fact that critical mass is now available in nontraditional literature. These grass roots or homegrown systems of dissemination and archiving have become powerful ways to gain recognition.

Economics is the power of persuasion. When libraries are engaged in more cancellation projects and reluctant to renew holdings than ever before, this should tell publishers that something must be changed. Grey literature does not have an easy crusade ahead. Technology clearly poses problems for the journal. At my own institution today, while we are here in Amsterdam, faculty colleagues and librarians are holding a seminar in Southern California to entertain ways they can influence the academic review system. The goal is to take a kinder more responsive (and more responsible) attitude toward rewarding scholars to contribute to highly regarded peer-reviewed open access models where it is not expected that the only way to publish is to turn over one's copyright. Today, authors find that no longer acceptable. Again, we reinforce that the current economic models cannot be sustained and that we need additional solutions to the unresolved problems of archiving and future reading as computing and technology continue to evolve.

In conclusion, the roadmap of grey literature is a series of bumps that have been smoothed over by new paradigms in technology usage and applications. The more ubiquitous a practice becomes the less parochial the interests are and the more innovation is practiced. As Raghuraman writes, "The writing is on the wall: Scientists also should think like engineers and engineers also should think like scientists."<sup>10</sup> It is also clearer to see today how when Mark Weiser coined the term, ubiquitous computing, we were not sure what the scaling of being surrounded by computers everywhere would involve. Not only are we using desktops, traveling with computers, carrying them all the time, wearing them, they are totally embedded in our world. Thus, we are challenged to see how we will design, use, and understand the elements of a ubiquitous computing world and our interest in grey literature is just one micro application. Weiser wanted ubiquitous computing to become invisible, but saw the need for better design for context awareness and that has become true for grey literature. We must continue to ask questions, entertain new possibilities and collaborate with a range of specialists to devise and plan new applications and products. There will certainly be more grey, but will there be less black and white as a result? Personally, I have to conclude that grey literature will have a new home in the spectrum of digital libraries. The new repository systems that are being hosted by a

variety of institutions where there is a welcoming "stewardship" as Clifford Lynch refers to it. The homecoming and lifespan of this new content in a connected context may be shorter but the diversity may be richer and "new analysis environments" will evolve and new communities of collaborators may result. As Clifford Lynch and Charles Henry recently ponder the future of digital libraries in interviews conducted for *Ubiquity*, Henry suggests, "something very different will emerge - something that is shared, collaborative, that doesn't belong to you or to me or to a particular institution, but belongs to everyone and no one perhaps at the same time. This is where a lot of academic information will reside 50 years from now. It will be very interesting to see how an institution navigates through these contradictory impulses."<sup>11</sup> There is a lot more to dissect, understand and grey literature will survive this dizzy roller coaster but perhaps in a different shade of grey, with a different price tag and different access architecture surrounding it. One of the most remarkable works that revolutionizes design and thinking about all this abstracting in a most clear way is the book by Paul Dourish, *Where the Action Is: The Foundations of Embodied Interaction*. If all creators of grey literature would follow the HCI design lessons noted, life in a wired world would be far easier and the publishing and library industries would be more advanced. Difficult yet pragmatic, a phenomenological approach such as what Dourish offers with an understanding of "how social practice emerges, we can build systems that fit more easily into the ways in which we work and live."<sup>12</sup>

---

<sup>1</sup> Marcia Riley, "Ubiquitous Computing: An Interesting New Paradigm,"

[http://www.cc.gatech.edu/classes/cs6751\\_97\\_fall/projects/say-cheese/marcia/mfinal.html](http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/say-cheese/marcia/mfinal.html)

<sup>2</sup> John Barell, *Developing More Curious Minds*. Alexandria, VA: Association for Supervision and Curriculum Development, 2003: 18.

<sup>3</sup> *Ibid*: 163.

<sup>4</sup> Scott Montgomery, *The Chicago Guide to Communicating Science*. Chicago, IL: University of Chicago Press, 2003: 187.

<sup>5</sup> Interview with Richard Leifer on Radical Innovation in *Ubiquity*, v. 1, Issue 47, February 6-12, 2001.

[http://www.acm.org/ubiquity/interviews/r\\_leifer\\_1.html](http://www.acm.org/ubiquity/interviews/r_leifer_1.html)

<sup>6</sup> Interview with Esther Dyson, "Promoting the Possible," in *Ubiquity*, vol. 2, issue 1, February 20, 2001.

[http://www.acm.org/ubiquity/interviews/e\\_dyson\\_1.html](http://www.acm.org/ubiquity/interviews/e_dyson_1.html)

<sup>7</sup> Francis Cairncross, "Knowledge, Decision-Making and Innovation," in *The Company of the Future: How the Communications Revolution is Changing Management*, chapter 2. Boston, MA: Harvard Business School Press, 2002:2-3. Reprinted in *Ubiquity*, vol. 3, Issue 1, February 19-25, 2002.

<sup>8</sup> *Ibid*: 19.

<sup>9</sup> Anne Galloway, "Resonances and Everyday Life: Ubiquitous Computing and the City."

[http://www.purselipsquarejaw.org/mobile/cult\\_studies\\_draft.html](http://www.purselipsquarejaw.org/mobile/cult_studies_draft.html)

<sup>10</sup> R. Raghuraman, "Now Hiring: Sciengineer," *Ubiquity*, vol. 1, Issue 43, 2001.

[http://www.acm.org/ubiquity/views/r\\_raghuraman\\_3.html](http://www.acm.org/ubiquity/views/r_raghuraman_3.html)

<sup>11</sup> Charles Henry, "Redefining the Role of the Library," interview in *Ubiquity*, vol. 4, Issue 25, August 2003.

<sup>12</sup> Paul Dourish, *Where the Action is: The Foundations of Embodied Interaction*. Cambridge, MA: MIT Press, 2001, from <http://www.dourish.com/embodied>

---

## Author Information

### Julia Gelfand

Science Library 228, Research & Instructional Services

University of California, Irvine (UCI)

P.O. Box 19556

Irvine, CA 92623-9556

USA

[jgelfand@uci.edu](mailto:jgelfand@uci.edu)

[www.uci.edu](http://www.uci.edu)



## **Grisemine, a digital library of grey university literature**

Marie-France CLAEREBOUT  
Université des sciences et technologies de Lille

### **ABSTRACT**

The progressive change from printed to digital support, induced by the policy of the commercial publishers, is likely to reduce little by little our university library to being a passive relay, and consequently to question our existence in the circuit of the dissemination of scientific information.

Our survival depends on our capacity of adaptation. That is why we chose to create without delay our own digital library of grey literature, Grisemine, and to transpose our know-how as information specialists in the necessary tasks, which are indexing, cataloguing and organization of the access to documents.

Created at the end of 2001, Grisemine has gradually grown richer by a selection of teaching or research documents: theses and dissertations, courses, communications, scientific reports... in the various disciplines, which our university library covers.

The favorable reaction of the contacted authors, as well as the increasing frequentation of our site, (<http://bibliotheques.univ-lille1.fr/grisemine>), confirm the relevance of our step.

It still needs to reach a critical size, a principal guarantee of credibility. We hope to reach that point while developing the co-operation with other actors engaged in the production and/or dissemination of grey university literature.

### **THE CONTEXT**

For years, our university library has been used to collecting printed grey literature and communicating it to her readers. Most of those papers (about 60,000) are theses and come from national or international exchange among libraries. The others are given by institutions or by authors themselves: working papers, conference proceedings, official reports, duplicated lecture notes, exam subjects...

#### **In a moving world, ...**

But, in the same time, dissemination of grey literature has become a stake for researchers, in reaction to the prices inflation of commercial publications.

That is why we can witness an evolution of the scholarly information dissemination, initiated by the authors themselves and helped by the development of the World Wide Web. Before being accepted by well-known journals, articles are quickly proposed to the community on their author's home page, on the institution's pages or in some specialized preprint open archives. Indeed, the increasing possibilities brought about by electronics allow a new way of dissemination for knowledge, the Net. That is the way chosen by commercial publishers; subscribers of most scientific journals can directly read them from their laboratories, on their computer screens.

Does it mean that while knowledge used to be disseminated by means of libraries within the context of printed publications, it will be disseminated by means of network in the world of electronic information? Are new technologies going to replace libraries and librarians?

#### **...we have to transpose our skills**

We, academic libraries, have to worry about that: are we becoming less involved in the scholarly publishing process?

The response depends on us. Librarians have to show that taking in account their know-how will help the whole community.

A first step is going on providing information itself. Our traditional role still exists, however documents are in printed or electronic form, and this role is to collect, catalog, preserve, share, provide easy access to texts.

As the physical library is a central part of the university, the digital library ought to be a central part of the “on-line university”, the non-for-profit actor offering necessary documentary services as a supplement to educational services.

Besides providing access to more or less expensive e-journals, our website should offer access to grey literature where an important sum of knowledge lies. Our confirmed skills allow us to build by ourselves organized grey literature repositories; I mean true “digital libraries”, which we have started to do with Grisemine.

## **A FIRST STEP: GRISEMINE**

At the end of 2001, we created Grisemine, a digital library destined to be a mine of grey literature. Grisemine is gradually growing, slowly because of our limited means but safely because of the value of its contents.

Several goals are pursued:

- giving more visibility to French-written scholarly production
- bringing our usual readers, scientists and students, wider access to research
- developing new collections
- getting new distant readers
- developing an economical complement to commercial scholarly publishing

### **Contents on line**

In this first step, we only collect French grey literature. A development plan was defined: aiming at a relative weight for each selected discipline and for each selected type of document.

How do we select?

Texts from any country and in any discipline covered by our university are accepted: sciences, technology, and social sciences.

We look for three types of documents, produced by universities, research or teaching institutes, public organizations:

- First, research texts which have already been communicated: theses and dissertations, symposium communications, articles of scholarly journals (journals published by societies, many of which do not yet provide electronic versions of their journals)
- Second, research texts which have not yet been communicated: notes, working papers, preprints
- Third, teaching texts: courses, exercises, exams

### **Functionalities**

Like a traditional library, with Grisemine we have to collect, catalog, preserve texts and make each text easy to retrieve

We try to answer our readers’ current needs, using today’s tools. So far we have kept a hybrid approach: most resources are both available in printed form in the physical library and in digital form on the network.

We apply new technologies for making the full text available in electronic format through a search interface on the Internet

A user can either

- browse the digital library by subject or by discipline
- search across title, author’s name, keywords, table of contents, ...
- search inside full text
- freely print or download texts

So we offer distant printing and navigation inside the text, which are not actually traditional services from a library.

### **Scientific and legal aspects**

One of the most frequently asked question is “Some texts are already available on other Websites. So, why do you not simply provide links to those sites?”

It could be an approach, but ours is quite different: as librarians, we feel that our mission is to provide faithful, high quality and secure contents.

Asking for a copy of each text, and building our own library, can seem heavy. However, controlling all the chain is also a guarantee of durability: we cannot risk losing precious documents, just on account of a broken link.

It is a guarantee of credibility and quality, too. We can refuse to communicate a text that would not be considered as scientifically sure. For any writer to cooperate with Grisemine, the only requirement is to belong to a known research or teaching institution. So, with just a little more control, we are able to provide rapid dissemination. But, if needed, a peer review could easily be envisaged.

Before any copying or dissemination, we ask the copyright holder for rights. Sad notice: in the case of previous (even non commercial) publication, authors too often lose control of the intellectual property or, at least, are not sure they have kept it.

### **Technical aspects**

We collect French grey literature, in both printed and electronic formats. By now, most of the documents exist in primary digital sources, which helps to provide access in a cheap way.

For insuring security and fair use of the electronic resources, texts are converted in PDF format, protected against ‘copy-paste’. So the distant-printed version matches the appearance of the original one. A system of bookmarks enables the use of the table of contents as a navigation tool inside the text, while the hidden text permits to search inside the full text.

We are planning to offer HTML format too, with the authors' agreements.

### **Bibliographic description**

The more documents you get, the more structuring you need, whether the format is printed or electronic. Our search interface on the Internet<sup>1</sup>, reflects our tools and methods in bibliographic description:

Each electronic text is linked to metadata tagged in MarcXML. Created by the Library of Congress, the MarcXML format mixes strictness of Marc, the librarian's format, with interoperability offered by XML.

Another guarantee for interoperability is to maintain international standards inside those metadata fields. Each text has a double discipline classification: a local one, useful for searching on the Website, and an international one, the Dewey classification, useful for metadata exchange or pooling.

With the same objectives, which are pertinence of results on the Website and interoperability, we also systematically define normalized keywords beside those chosen by the authors.

For offering more pertinent search fields, the description of each text also includes the abstract and table of contents. It also includes the original quotation if the text has been loaded from another Website.

---

<sup>1</sup> <http://bibliotheques.univ-lille1.fr/grisemine>

## PROSPECTS

### Grisemine's feedback

After two years, results have been quite good: in spite of the small quantity of available texts (about 750 so far), frequentation of our website is steadily increasing.

Originally, we had to communicate a lot for being known. Now, people come more spontaneously to our site. And these are probably not only French-speaking people: from the beginning of 2003, visitors have come from 74 different countries, which encourages us to collect and disseminate in other languages which are too scarce on the Web...

Who are our distant readers and what are they looking for ?

A predictable noticing: they do not only come from institutions. Several firms, very interested in scholarly grey literature, asking for texts on some specific issues, have contacted us.

We have also noticed a great interest for courses: although they represent only 11% of the contents, courses account for over 30% of accesses. So, on the last revision of our development plan, we decided to increase the proportion of teaching texts inside the digital library.

And what about the authors ?

When contacted, authors usually agree to give us their texts (even if they too often forget to, and have to be re-contacted...).

A few new authors have begun to willingly send their texts, too. Their trust in the visibility offered by our library is a very good credibility indicator.

### Closer to the authors

A new challenge for us, librarians, will be to take in account the structuring of contents. Even if grey literature is heterogeneous in essence, some kinds of documents can be made homogeneous from the creation: theses and dissertations, communications, articles. What should we ask the author to respect a few structuring rules for?

To get a better printed version, as some publishers or symposium organizers do. The aim is also one of preservation, too, because a structured document can be archived in a durable format. And, why not, in order to allow further modular re-use of the text.

In France, our university is one of the first ones engaged in the process of structuring theses for electronic diffusion and conservation: they will be tagged for XML encoding, designed to identify each element of the title page and each part of the thesis. It is with this aim in view that the library gives the authors all useful recommendations about the style sheet to use, as soon as they begin to write.

Another way of development for university libraries will be the creation of institutional repositories. Indeed, more and more institutions want to keep ownership of information produced by their members and so, they plan to create a digital repository for their entire intellectual output. This way, libraries can be good services providers.

Those two new aims add to our role of information providers a more service-oriented role and lead us to a new position in the information cycle of information creation, distribution and use. Quite a radical change...

### Collaborating with other actors

In fact, Grisemine can be seen as a prototype over other projects of scholarly grey literature dissemination, often led by researchers themselves. There still remains to manage vast amounts of electronic information, easy to retrieve, which lies on cooperation among the different actors.

Generally, all not-for-profit actors would probably take advantage of cooperation between each other: libraries, higher education, researchers, scientific societies, not-for-profit publishers, university presses,...

First of all, documents providers should cooperate by exchanging full texts or metadata only, so they would be able to constitute large libraries or large metadata repositories. Grisemine will soon be able to do this by becoming OAI-compliant, which means that it will respect the existing Open Archive Initiative protocol<sup>2</sup>, accompanied with still missing normalized semantic rules.

### Author Information

---

**Ms. Marie-France Claerebout**, Librarian  
Université des sciences et technologies de Lille (Lille 1)  
BP 155  
59653 Villeneuve d'Ascq cedex  
France

Marie-France.Claerebout@univ-lille1.fr  
<http://bibliotheques.univ-lille1.fr/grisemine>

---

<sup>2</sup> <http://www.openarchives.org>

## Quality of grey literature in the open access era: Privilege and Responsibility

Paola De Castro, Sandra Salinetti  
Istituto Superiore di Sanità, Rome (Italy)

### Abstract

The privilege recently gained in the dissemination of Grey Literature (GL) through the widespread use of the Internet implies a new level of responsibility for GL authors and issuing organizations. In the past, in fact, only a limited number of copies were produced, generally in-house printed, and the target readers were very well focused and known beforehand. Now in the open access era, greater care should be assured not only to the scientific content of GL, but also to the respect of editorial standards to allow easy readability, understanding and retrieval of the document in the Internet. The differences in the editorial processing of GL documents *versus* articles published in open literature stress the uncertain borders existing between the two channels of publication. In particular, the different levels of technical editing, applied also to non-officially refereed documents, can be compared to the editorial process followed in conventional literature. The long experience in the production of GL by the Istituto Superiore di Sanità (the Italian National Institute of Health) allows reflecting on these issues related to GL production, circulation and use. Besides the authors' primary responsibilities, it is important to focus on the educational role of the editorial staff (inside the research institute producing also open literature), actively contributing to guarantee a better quality to all documents issued by the organization.

For correspondence: [paola.decastro@iss.it](mailto:paola.decastro@iss.it)

### The grey nuances: from past to present

The concept of Grey Literature (GL) has suffered many changes since the Seminar of York (1978) which set the basis for the discussion on this kind of material in Europe.<sup>1</sup>

At present, the awareness of the specific role of GL, within the information transfer process, is no longer questioned. Its importance and usefulness is now widely recognised, whereas in York it was even difficult to find a proper definition for GL. A simple search, through Google, for the term "grey literature" (that once was known only to a small group of specialists and was called with different terms such as report or fugitive literature) gives more than 600,000 scores.

Moreover, today the Internet revolution has affected also GL, while in the First International Conference on GL, held in Amsterdam in 1993, network publishing represented just one of the many issues at stake.<sup>2</sup>

Yet, some of the characteristics historically associated to this literature are still valid, in particular:

- *Issuing organizations*  
They continue to play a fundamental role in the production and dissemination of GL, mainly in the scientific field.
- *Document types*

---

<sup>1</sup> Alberani V, De Castro P. Grey literature from the York Seminar (UK) of 1978 to the year 2000. *Inspec* 2001;35(4):236-47.

<sup>2</sup> First International Conference on Grey Literature. Rai Congress Centre. Amsterdam, The Netherlands. December 13-15, 1993. Conference Proceedings. Amsterdam: *Transatlantic*, 1994.

Types traditionally included in GL are still produced (technical reports, conference proceedings, translations, theses, etc.), even if new types of documents are added to the traditional ones (e-prints that have taken the place of pre-prints, collections of FAQs, some e-journals, personal pages of individual researchers, power point presentations<sup>3</sup>, etc.).

– *Value/costs*

GL intrinsic value and its production costs remain bound in an inverse relationship since the funds to carry out the activity described in GL far exceed production and diffusion expenses.

– *Presentation*

The importance of a correct presentation continues to have a fundamental function as it is directly associated to document understanding, retrieval and availability, even if Internet sources are mostly used and documents position in the web deeply influence their traceability.

– *No referee process*

GL is still not subject to the peer-review process, differently from conventional literature.

The role of the Net was not predictable in the two cited “historical” meetings: the bulk of information produced is now dramatically increasing and most documents are available online regardless of their type (open or grey literature) and especially of their quality.

The Internet, in fact, deeply transformed some GL features as mainly regards:

– *Information retrieval*

Library catalogues and other information sources about GL documents or producers are available in the Internet at institutional, governmental or inter-governmental levels or in specific subject areas, for example: the British Library, with a long tradition in GL supply, has now its online free catalogue<sup>4</sup>; the National Library of Medicine, through Locator plus (the online catalogue of its collection), provides bibliographic information on many report series in the biomedical field<sup>5</sup>; the New York Academy of Medicine offers a list of US Non Profit Organizations and Government Agencies producing GL in public health and health policies<sup>6</sup>.

– *Document availability*

Access to GL full-text documents is easier now, the document URL helps more than the report number to get the document (if a site is properly organized and maintained).

– *Awareness*

A growing interest for GL is expressed by professionals both in the field of librarianship and information science (e.g., the National Library of Medicine includes a chapter on GL in the health technology assessment information resources<sup>7</sup>), and in other specialized fields recognising the importance of documents created for practical more than prestige purposes (e.g., GL now appears as a fundamental source also in meta-analysis of randomised clinical trials, especially when negative results are concerned<sup>8,9</sup>).

Moreover, online circulation of electronic documents has been adding confusion to the overall editorial scene: GL documents are more and more comparable to other types of

<sup>3</sup> A good example of an online collection of power point presentations in epidemiology and related fields is available at: <http://www.pitt.edu/~super1/index.htm>.

<sup>4</sup> <http://www.blpc.bl.uk>.

<sup>5</sup> <http://www.locatorplus.gov>.

<sup>6</sup> <http://www.nyam.org/library/greylit/greylitorgs.shtml>

<sup>7</sup> [http://www.nlm.nih.gov/nichsr/ehta/chapter\\_10.html](http://www.nlm.nih.gov/nichsr/ehta/chapter_10.html).

<sup>8</sup> Hopewell S, Mc Donald S, Clarke M, Egger M Grey literature in meta-analyses of randomized trials of health care interventions. *The Cochrane library* 3: 2003.

<sup>9</sup> McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature estimates of intervention effectiveness reported in meta-analyses? *The Lancet* 2000;356:1228-31.

traditional documents, for example: the documents deposited in institutional archives are not or not yet subject to the editors' approval;<sup>10</sup> the layout of GL documents is not so "grey" as it was in the past; at first glance, there is no evident difference between grey or white electronic journals; etc.

Crisis or change applies also to conventional literature. In 1995, Roland LaPorte published a meaningful article in the *BMJ* stating the death of scientific journals<sup>11</sup>; some years later, in 2002, the same author reconsidered the concept of death and changed it into metamorphosis that seemed to be more appropriate for this new condition<sup>12</sup>. In fact, scientific journals are changing their original shape as well as GL, but nobody can predict the future. Paraphrasing LaPorte, may it be that also the grey bug of GL is transforming itself into a colourful butterfly?

### The Internet "sea change": where is the privilege?

The privilege accruing to GL producers and users thanks to the easy diffusion of GL through the Internet is clear: once the document is "ready", you can disseminate it through the Net without much effort and at very low costs.

As regards producers, some of the advantages of including GL in their organization websites are:

- duplication costs or shipping charges are reduced to zero;
- publication run is no longer a big trouble;
- online GL is never out of stock and never to be photocopied;
- deposit spaces for extra copies are no longer an additional burden;
- online GL may be easily linked with other documents, thus further improving its circulation;
- etc.

From the users' point of view, the benefits are:

- direct availability of GL (once you know the name of the issuing organization);
- access to other GL documents which might result from different types of searches (through searching engines, databases, etc.);
- exploitation of hyperlinks to other relevant documentation (be it grey or not);
- immediate feedback (reader-author-reader) through e-mail contact;
- etc.

All this is true, but the Internet sometimes creates illusions. Most GL documents are now on the Net, however, it is undeniable that their access is not always so easy. There is a tremendous risk of puzzling many Internet users and misleading potential readers when the source and target of documents cannot be clearly identified, that is if documents are not placed in the correct position on the institutional servers or if they are not properly conceived and structured. When you find the correct URL, pointing to a specific document, there should not be obstacles to read, print, or save it on your PC. Yet, URLs may change in time and document traceability may still be a difficult task.

### The Internet challenge: where does the responsibility lay?

The privilege gained in the dissemination of GL through the widespread use of the Internet implies a new level of responsibility for GL authors and issuing organizations.

In the past, GL was produced only in a limited number of copies, generally in-house printed, and the target readers (experts in specific fields) were well focused and known beforehand. The information contained in GL was not always formally correct (i.e., well

<sup>10</sup> Guedon JC. In Oldenburg's long shadow: librarians, research scientists, publishers, and the control of scientific publishing. Available at: <http://www.arl.org/arl/proceedings/138/guedon.html>.

<sup>11</sup> LaPorte RE, Akazawa S, Gamboa C, Maclure M, Marler E, Sauer F, Shenton C. The death of biomedical journals. *BMJ* 1995; 310: 1387-90. Available at: <http://bmj.com/cgi/content/full/310/6991/1387>.

<sup>12</sup> LaPorte RE, Linkov F, Villasenor T, Sauer F, Gamboa C, Lovalekar M, Shubnikov E, Sekikawa A, Sa E. Papyrus to PowerPoint (P2P): Metamorphosis of scientific communication. *BMJ* 2002;325:1478-81.



structured, containing all relevant bibliographic data, complete references, etc.), but since the target was clearly identified, the possibility of misinterpretation was much rarer than it is today when also GL is part of the wider category of open access documents.

Dissemination at low costs does not solve all problems: from an editorial point of view, commitment in the production of GL has increased since the Internet does not “forgive” any mistake and the whole world may look at you!

The problem is that, traditionally, GL has never been subject to any referee control as it was not to be published. Its aim was that authors could receive an immediate feedback and target readers could quickly use information. Now the term “publish” has acquired a different meaning and publishing GL in the Internet is an accepted concept, but requires a different editorial approach.

GL remains a very rich and unique source of information that would seldom appear as such in the open literature (technical reports contain very detailed information, many tables or figures that a scientific journal would never accept for publication, etc.). Before spreading it in the Internet, therefore, it must be properly evaluated, also because it is no longer possible to have a complete control of the target (for example, the general public having access to specialized GL medical information may misuse specific prescriptions addressed only to medical staff).

### **Why does editorial quality become so relevant in the open access era?**

Now that access to GL may be theoretically guaranteed to a worldwide audience, greater care should be assured not only to the scientific content (that was, is and will always be the most relevant feature in the evaluation of any paper), but also to the respect of editorial standards and rules allowing easy understanding and retrieval of the document itself.

Until the end of the 1970s, there was little concern in Europe about the correct presentation of GL, which was often lacking the basic elements for its identification, not to consider its structure and presentation.

The Seminar of York in 1978 produced a specific recommendation to improve GL presentation and therefore its circulation. As a result, the ISO Standard 5966 on the presentation of technical reports was issued in 1982.<sup>13</sup> This standard is considered the first step towards a better production and exploitation of technical reports and similar material (thesis, translations, etc.) satisfying both librarians' requirements and basic editorial rules.

The use of this standard among GL producers was highly recommended and promoted at different levels by librarians, documentalists and producers of GL (it was also translated into Italian and published by the Italian standardization authority, UNI<sup>14</sup>). Thanks to this collaborative effort, many documents are now produced with much greater care, thus guaranteeing their basic bibliographic control.

As already mentioned, however, the standard ISO 5966 was issued in 1982 and many things changed since then. In fact, it was withdrawn by ISO and now no trace of it is left in ISO online catalogue (available at: [www.iso.ch](http://www.iso.ch)).

Just to quote an example from this standard as regards mathematical, physical and chemical formulae, it states:

“The presentation of mathematics shall follow conventional practices. It is, therefore, often preferable to use careful handwriting rather than to attempt to reproduce mathematical notations on a conventional typewriter. However, when special printing, typewriting, or drafting facilities are available for the preparation of mathematical notations, these should always be used.”

Today, this statement (and other recommendations contained in the cited standard) just makes us smile.

<sup>13</sup> International Organization for Standardization. *Presentation of scientific and technical reports*. Geneva: ISO; 1982.

<sup>14</sup> UNI ISO 5966. *Presentazione di rapporti tecnici e scientifici*. Milano: Ente Nazionale Italiano di Unificazione (UNI); 1989. (Italian translation of ISO 5966).

The reasons why the standard was withdrawn are easy to explain as dramatic changes occurred in the production and diffusion of GL since 1982; however, the basic philosophy regulating it still applies even in the Internet times, for example the following items should not be disregarded:

- elements that should appear on the title page of a technical report (issuing organization, authors' names and affiliations, title and date of publication, ISSN or/and ISBN, report identifier, etc.);
- abstract;
- document structure (introduction, core of report, conclusion);
- report numbering;
- indexes;
- quantities, units and symbols;
- etc.

The widespread use of word processing programs helps a lot, but it has not solved all the difficulties encountered by GL authors.

An effective strategy to improve GL quality is pointing to the empowerment of authors through specific training courses on scientific writing.

In the last 20 years, we have been witnessing to this phenomenon: thanks to the support of the new technologies, authors seem to be entitled to become editors of their papers (both grey and not). Actually, they have acquired new skills in using word processing programs, but still need basic training in editorial principles to be qualified producers of documents. Technologies help to perform some tasks, but you must be able to exploit them correctly: a technical document might be typographically perfect, have excellent design and layout, no grammatical errors but be pointless if the information therein contained is not consistent and properly organized. Appearance alone may be misleading!

Editorially speaking, there are many strategies or rules to guarantee added value to a high-content document. In some cases, it may be useful to organize a report into sections and subsections (either with numbers or different character styles or sizes). Widely used software, such as "Word" (that is now the standard in word processing, both in IBM or Mac platforms), may give hints to authors, but they must be aware of the hierarchical division of their ideas and of the importance to structure their texts. Readers must not be disorientated in their approach to a document, either printed or electronic.

A good structure means readability and ease of use, and it also helps to retain information.

Traditionally, GL has not been not subject to the referees' judgement, but it is obvious that a certain type of revision is generally applied within an organization before its wide diffusion or online publication.

All types of documents should be editorially revised before being issued.

The Council of Biology Editors (now Council of Science Editors) has envisaged three levels of revision<sup>15</sup> depending on: type of document, time for revision, budget, circulation, etc. The purpose is that of ensuring: consistency (that is preventing contradictory statements in the abstract, introduction or conclusion); completeness (for example, that all tables and figures, cited in the text, must be present); accuracy (in the use of symbols, acronyms, units of measurements, or references, etc.); clear understanding to the intended audience; etc.

The three levels of technical editing (rush, standard, and revision editing) consider the same basic revision principles but with different degrees of attention to details.

Rush edit is compulsory even in GL, since it implies policy, copyright, copyediting and technical content considerations that can never be disregarded even when circulation is limited. When time or budget allow the application of more careful revision (standard and revision editing, which take more specifically into account integrity, style, language, accuracy in references, etc.), the document will be further improved for the sake of all readers and the image of the issuing organization.

GL authors should be aware of these editorial rules even more than authors of documents to be published in open literature; they will not benefit, in fact, from the referees' or editors'

---

<sup>15</sup> David E. Nadziejka. *Levels of technical editing*. Reston (Virginia): Council of Biology Editors; 1999. (Guidelines 4).

contribution giving added value to their original paper in the different stages of the editorial process.

In any scientific field, researchers are used to ask for the informal consent or approval of the closest colleagues before spreading a document. A more specific internal review process for GL is sometimes required within a scientific structure, for example the supervision of the Head of Department or Project Leader, or other scientific committees within the organization. Therefore, even if there is no formal peer review, "silent" levels of revision are applied and highly recommended, especially when GL is available online.

Editorial quality is also directly connected to document retrieval.

In the past, librarians and documentalists had problems in retrieving and having access to GL circulating only through the "invisible colleges" as it too often lacked the basic elements (authors, title, issuing organization, publication date, etc.) permitting its bibliographic control.

Moreover, GL producers generally rejected the idea of making a widespread dissemination of their print production, but seldom denied the inclusion of GL in specialized databases, thus promoting the diffusion of this material in the interest of the scientific community (the target, however, remained limited to small groups, due to the specialized subjects covered).

As a general rule, bibliographic control was and still is possible only when a document contains those elements allowing both its processing in bibliographic databases and its correct citation in other publications. Also GL producers should be aware of this responsibility in issuing documents.

Attention to the title has also a primary role in assuring a quality document. To guarantee GL retrieval, title words must be properly selected to allow their exploitation in bibliographic databases. For this reason, also the role of descriptors or keywords should not be disregarded.

In fact, an editorially correct document (with appropriate title and descriptors, solid and consistent structure, etc.) always enhances indexing in existing databases and improves its retrieval through Internet searching engines and other facilities, as well.

## The experience of the Istituto Superiore di Sanità

The long dated editorial role of the Istituto Superiore di Sanità (ISS, the Italian National Institute of Health) allows reflecting more closely on the old and new privileges and responsibilities of GL issuing organizations. The ISS has an editorial service producing an official journal (*Annali dell'Istituto Superiore di Sanità*, printed since 1936) and different types of GL (such as *Rapporti ISTISAN*, the main series of technical and research reports started in 1977, and *Notiziario dell'Istituto Superiore di Sanità*, a monthly newsletter produced since 1988).

The most abundant production of printed GL was in the 1980s-1990s, before the widespread use of the electronic medium<sup>16,17,18,19,20</sup>, due to the success of the informal communication channel in that period.

Later, the benefits acquired through the Internet made it useless to continue the production of some GL series, since researchers began to use other quick forms of communication (transfer of documents to small groups by e-mail attached files, discussion lists, etc.) thus creating new types of GL, once again far from any bibliographic control. PDF files of the ISS technical reports began to appear in the Institute's website since 1996, but, only in 2001, all the ISS production was made regularly available on the Net. This required more careful editorial control. Technical reports revision by the editorial staff became more and more careful through the years (instructions to authors are now very detailed and even a model file for writing reports is used for help authors formatting their texts). Therefore, the quality of reports (still produced in a limited number of paper copies) is now much higher than that of the reports produced only in hard copy until a few years ago. If you compare the layout of ISS non-commercial publications through the years, the differences are astonishing.

<sup>16</sup> *Aggiornamento bibliografico in igiene ambientale (1980-1993)*

<sup>17</sup> *Eurospin Quarterly (1985-1993)*,

<sup>18</sup> *Bollettino Epidemiologico Nazionale (1982-1992)*.

<sup>19</sup> *Serie relazioni (1979-2000)*.

<sup>20</sup> *ISTISAN Congressi (1988- )*.

The activity of the editorial service paradoxically increased since GL has been online. ISS authors have become aware of their new responsibility and demand the editorial collaboration and advice much more than in the past.

The editorial staff has acquired the role of tutor to help researchers better perform their editorial tasks. Moreover, regular training courses are organized in scientific writing which prove very useful both to authors (learning how to cope with editorial tasks) and internal/external editors (receiving better manuscripts).

Editorially speaking, attention now given to the ISS not commercially published literature is not so different from that regularly applied to the quarterly journal *Annali*, except for the fact that the journal articles are subject to external peer review process while GL is revised only within the organization.

Furthermore, as regards GL typographical quality, the ISS reports and newsletter, that were previously in-house printed, are now produced in out-sourcing. This permits to use colour in covers of technical reports and to have a completely coloured newsletter thus guaranteeing a better print at lower prices; the number of copies produced, however, is still limited.

Print-on-demand facilities might be helpful to keep the whole production cycle inside the ISS, but they require careful evaluation and, by the moment, an outsource service seems to be the easier way to replace old printing machines.

*Rapporti ISTISAN* has always represented a valuable source of information, and now, thanks to the editorial contribution, it is more similar to “commercial monographs” than to GL. Also the newsletter, initially containing mainly news and practical information (for a quick circulation among health operators), now includes original articles or activity reports and the real “news” is just a small part of it. The newsletter look, now colour printed, is closer to a journal than to a grey document.

The borders between grey and not grey are getting more and more uncertain.

As regards retrieval of the ISS documents, the official journal is regularly indexed in Medline (the most important database for medical literature); therefore there is no obstacle in identifying any article therein published. Yet, the institutional GL production may still be difficult to get by those who do not know the Institute and searching engines do not always guarantee immediate access to the most relevant information. Once you learn the correct approach to the ISS production on the web, however, document exploitation is maximised. In particular, it is very much used at the peripheral levels of the national health system (health operators in the field of epidemiology, biotechnologies, environment, infectious diseases, etc.), in the private sector that could not otherwise have access to this documentation (that is not on sale) and in developing countries that have major problems also in receiving hard copies. Furthermore, all ISS reports are regularly included in SIGLE (System for Information on Grey Literature in Europe), and some of them are in NTIS (US National Technical Information Service), CAS (Chemical Abstracts Service), and other specialized databases. Only recently, they are available also through the Locator plus of the National Library of Medicine and other online catalogues at national and international level.

## Conclusions

We are living a deep change in the information transfer process involving the different actors of the editorial scene (from authors to editors, web-editors, e-publishers and readers), who occasionally play different roles at the same time (e.g., authors sometimes bypass the editor thus assuming direct responsibilities in the diffusion of documents in the Internet).

Furthermore, information producers, managers or seekers often show twofold personalities like Dr Jeckill and Mr Hyde: they would like to find all documents free in the web being sure to read original, reliable and high quality information, but they are sometimes reluctant to spread their papers through online non official channels as they care for impact factor, prestige and career.

It is a revolutionary period in which new and alternative forms of scholarly communication live together with more traditional ones and the future scenario is difficult to be defined.

Recent initiatives of open access - such as Pubmed Central, Public Library of Science, and Budapest Open Access Initiative - are contributing to define new trends in the editorial market, challenging the traditional distribution channels, mainly managed by commercial editors, and placing the authors' role and publication copyrights under severe discussion.

In this landscape, GL has now new dignity and becomes closer and closer to innovative scientific publications supported by researchers in view of a generalised movement towards open access. The nuances of grey are becoming lighter and lighter. The questions are many, the answers few, but, in any case, the ethical responsibility of producing and issuing quality documents can never be disregarded.

---

### **Author Information**

**Paola De Castro**  
Istituto Superiore di Sanità  
Viale Regina Elena, 299  
00161 Rome  
Italy

paola.decastro@iss.it  
Sandra.Salinetti@iss.it  
www.iss.it

## The communication flow of research projects results

Daniela Luzi, Rosa Di Cesare, Luciana Libutti  
CNR - Istituto di Ricerche sulla Popolazione e le Politiche Sociali, Roma

Maria Castriotta  
ISPESL, Dipartimento Documentazione, Informazione e Formazione, Roma

Mariarosaria Manco, LINK s.r.l., Roma

### Introduction

Research projects and scientific grey literature are linked by a cause-effect relation [1,2]. Scientific Grey Literature (GL) production is generally originated by research activities related to more or less formalised projects. Research projects, in their turn, are represented through GL documents (deliverables, technical reports, etc.) and it often happens that research products themselves are also GL (patents, protocols, guidelines, etc.).

This relation is underlined by two types of links pertaining to their *information content* and to their *process*. The first type consists in a set of information, which is common to the project description and to the GL documents produced within the project. The second type of link regards the process and, from this point of view, each information unit represents the result of a precise activity within the communication flow, carried out during the project lifecycle, from the drafting of the proposal, to its approval, until the finalisation of the research and the diffusion of the results.

The need to develop Research Information Systems (RIS) dates back to 1975, they were promoted by Unisist, the United Nations Educational, Scientific and Cultural Organisation [3]. The RIS, which at that time were simply called *registers of current research*, had two essential objectives, “to enhance communication among scientists concerning on-going projects” and “to provide an effective information base to managers of the national R&D program”. A simple search in the current CORDIS database [4] reveals a far more complex framework, not only in relation to the quantity of data, but above all for the services offered, which are developed to support both participation in European projects (lists of experts, deadlines for tenders, etc.) and their technological transfer (information targeted to companies and SMEs in particular).

As a matter of fact, the evolution of today’s CRIS (Current Research Information Systems) is at a crossroads [5] as they have become a research topic. On the one hand, they have to meet the needs of increasingly varied types of users (scientific community and policy makers, but also evaluators, intermediaries, companies, media); on the other hand they tend to represent the complexity of the world of research.

The trend is to build an integrated environment, which could represent and connect different types of information using the appropriate technology [6]. The various experiences related to the development of science portals [7, 8], or the systems automatically retrieving data from research web pages [9], national and international CRISs, all run in the direction of integration and interoperability. We too have considered this issue an important step and focused our attention on the integration of the information produced during the project lifecycle with its results, which is with the production of GL documents.

This paper describes the initial findings of a study carried out in collaboration between the Italian National Research Council (CNR) and the Italian National Institute of Safety and Health at Work (ISPESL) with the aim of developing an information system on research projects regarding occupational safety and health: RIS-OSH (Research Information System on Occupational Safety and Health). Such system will interface with the existing GL database (<http://www.ispesl.it/lg/default.htm>) and will allow the automatic updating of the documents produced within each research project. The ultimate goal of this commitment is to facilitate research evaluation activities and technological transfer through collection and updating of information related to research projects and to the documents produced within the process.

In this paper we will focus on the conceptual model of the RIS-OSH system and its integration with the GL database. The conceptual model is the result of the analysis carried out to obtain system requirements. The first paragraphs contain the rationale of the study and the

methods used. The results of the requirement analysis are reported in paragraphs 3, 4 and 5 respectively in the form of the workflow of a project; of the document transactions and GL production; and in the conceptual scheme of the integration between the information of the RIS-OSH and GL database.

## 1. Rationale

Information systems gathering research projects have always been quite different from other kinds of databases. Compared to bibliographic databases, for instance, RIS are distinct in that they “photograph” the lifecycle in the single phase of each project. The record describing the project is therefore progressively updated through the input of the information produced during the execution of the process according to precise time scheduling. This means that the effectiveness of the system is based less on the number of records fed into the database and more on the ability to “capture” reality both at the right time, and accurately.

In this way such systems, particularly if managed by funding organisations, can also be considered as monitoring tools of the whole process. For this reason, the hard task of updating a database becomes a crucial one. During the system design particular attention has to be put on monitoring the work steps providing tools for an easy input of the data to be updated by the identified actors.

The RIS-OSH system, which is at developmental stage aims to meet these requirements and can be considered within the framework of the so-called *corporate oriented* system [10]. Nevertheless, its design has taken the requirement of interoperability with other systems into account and it is therefore based on the CERIF 2000 standard [11]. The major entities of CERIF are {project, person, organisational unit} which are linked with the attributes {role}, {date/time}, {status}. These relationships allow flexibility and robustness in the description of the different environments in which research projects are carried out [12].

Generally, one of the weak points of CRISs is the diffusion of the results obtained during the project. The CERIF Task Group has put this issue on its research agenda, and foresees to extend the future CERIF 2000 standard to a more detailed description of research publications and products. Our study proposes an extension based on the integration of the data model related to ISPEL's projects with the data model, which describes GL documents. This solution makes it possible to provide different views and accesses to the information related to project results. One view embeds the description of the GL document in the context of the project in which GL has been produced; the other correlates the set of GL documents produced in different projects. The advantage of this model is twofold: the user can access a brief description of GL documents within RIS-OSH and, if interested retrieve a more detailed description in the GL database and/or access the full text of the document. From a bibliographic point of view GL documents can be described using a specialised classification scheme, which makes the correlation between the documents more evident. This is achieved through the use of metadata using XML to connect the different classifications and bibliographic schema [13].

## 2. Methods

System requirements have been obtained through the analysis of the process actually performed at ISPEL, when it funds research projects at external research institutions. This analysis has focused in particular on the workflow and the information flows featuring the different steps of the project lifecycle. The entire workflow has been classified into meaningful steps, and for each one, actors, tasks performed, time scheduling and exchange of information and documents have been identified. Documents produced during the process (Activity plan, Call for proposal, Contract) have been analysed on the basis of the information contained. This allows the reconstruction of the information flow as well as the transaction of the information derived from the work in progress.

The analysis was carried out through interviews to ISPEL privileged witnesses. In the preparation of interviews, particular attention was devoted to:

- The identification of tasks performed by two potential profiles of system users, head of department and researcher, in each step of the workflow, starting with the drafting of the proposal to the execution and conclusion of the project;
- The identification of their particular information needs targeted at developing information tools, which can help internal communication within each department and among departments, checking the consistency of data, searching and editing;
- The identification of the workflow steps and their related information that need an Intranet and controlled access as distinguished by those used for the diffusion of information on the Internet.

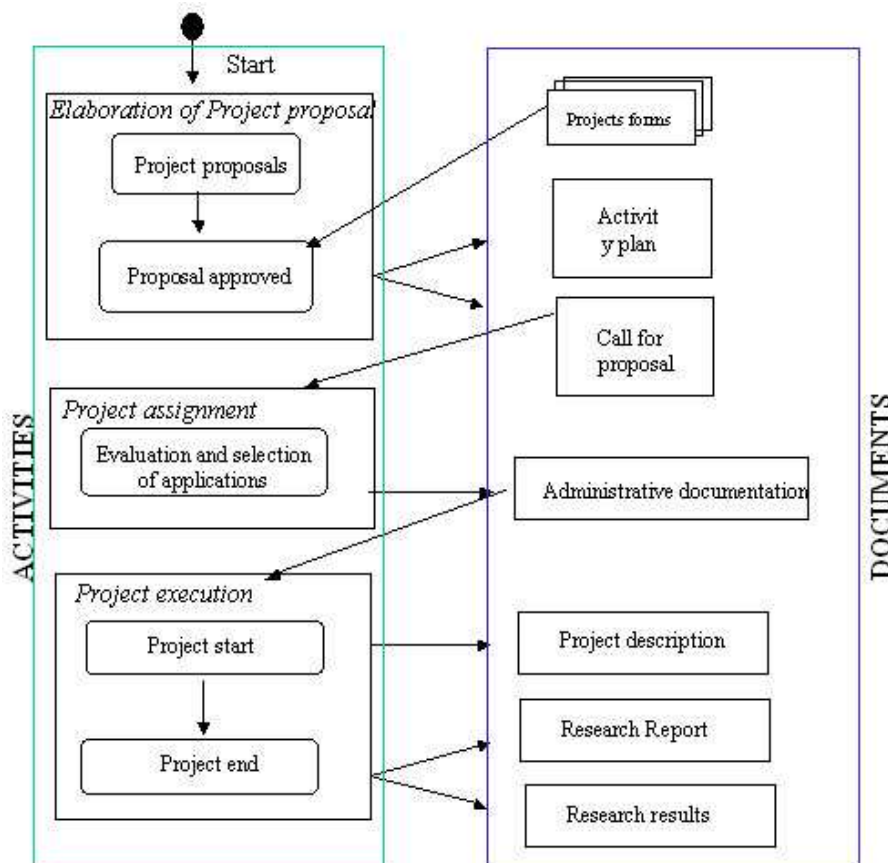
Through the interviews we could also ascertain the degree of acceptance to be involved in the future electronic system, as well as the availability of the potential users to participate in the experimental phase of the project.

We have used the standard UML (Unified Model Language) to describe the data and the workflow model [14].

## 2. The project lifecycle

The results of the interview made it possible to reconstruct the entire process carried out at ISPEL when it funds research projects at external institutions. Figure 1 summarises the main steps of the project lifecycle and shows the document produced in each step, while figures 2, 3 and 4 describe the process in a more detailed way, identifying the actors, the activities carried out and the document produced as a result of each activity.

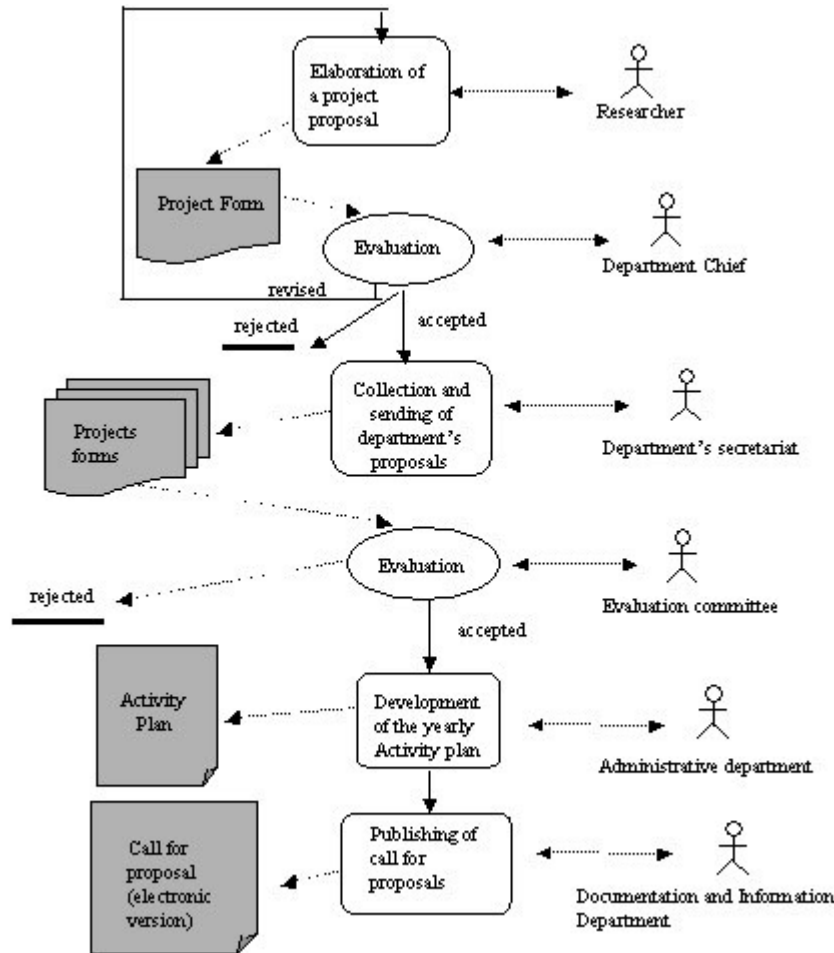
Figure 1: Main steps of the project lifecycle





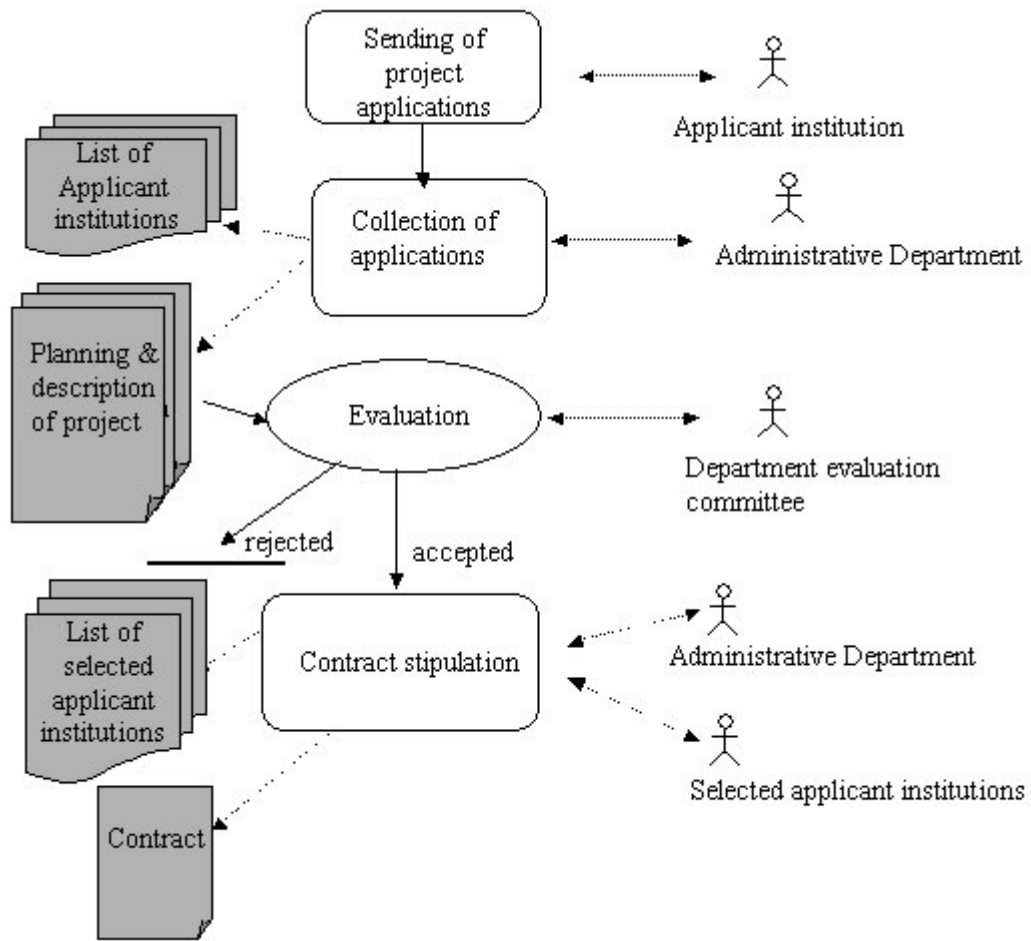
ISPESL’s funded projects are deeply integrated with its own research activities, they are inserted into pre-defined subject areas and research programmes, which are periodically established and revised, as they constitute the foundation of its mission. For this reason the ISPESL researcher, who elaborates a project proposal, supplies the potential applicant with a precise framework of the future project: a rationale and detailed objectives.

Figure 2: Elaboration and evaluation of project proposals



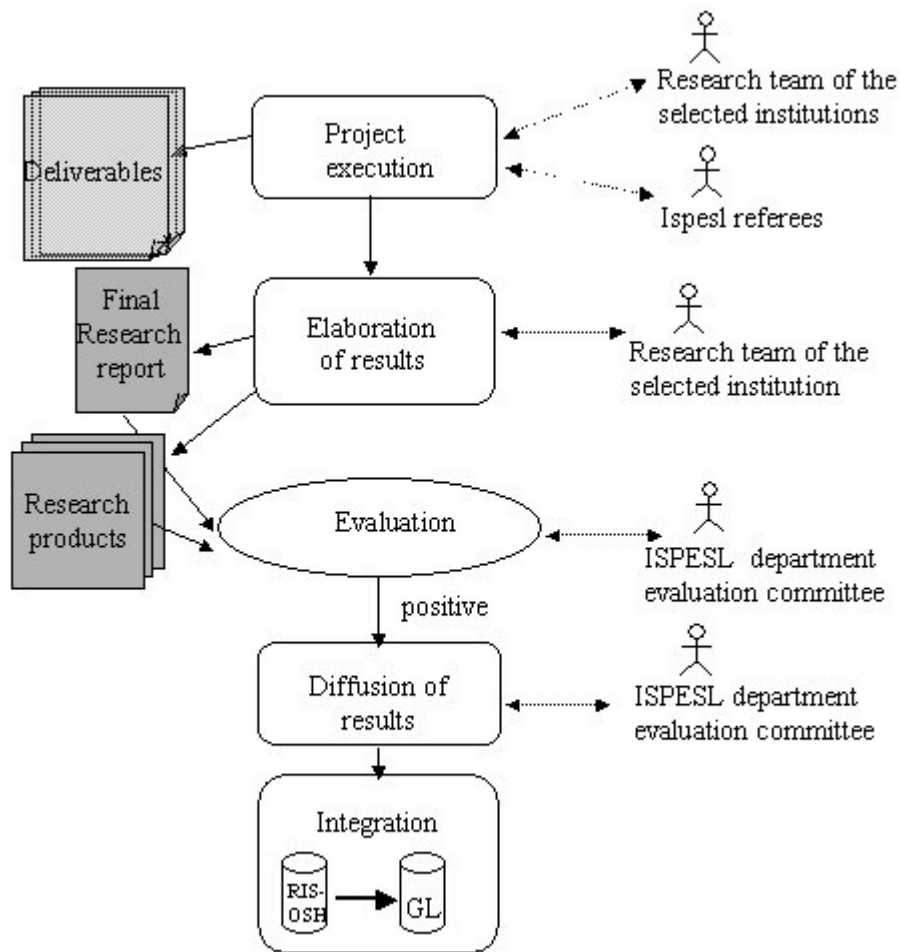
Project proposals are currently presented on a pre-printed form to the department chief, who evaluates the proposal according to the main objectives of the department, available funding and duration (fig.2). He may require the researcher to modify the proposal or he may even reject it. At a due date, each department collects and sends the proposals to the evaluation committee, where they are evaluated from an institutional, scientific and technical point of view. Once the proposals are approved, the Administrative department integrates them into the ISPESL yearly activity plan, while the Documentation and information department publishes the call for proposals on the Internet.

Figure 3: Project Assignment



At this point project assignment begins (fig.3): the external applicant institutions reply to the proposal sending their project planning and description to the administrative department, which circulates them for evaluation by the evaluation committee of each department. If the application is accepted, the administrative department stipulates a contract with the applicant research institution.

Figure 4: Project execution



Subsequently the project can get underway, and information about the ongoing project can be posted on the Internet to a larger target of users. Deliverables are produced to describe the progress of the project (this is the case for two and three years ISPESL projects) and the project is concluded when the final research report is delivered. The step of diffusion entails the integration of the RIS-OSH with the GL database.

### 3. The production of GL documents

The description of the project workflow has made it possible to identify the actors who participate in the process and the activities carried out, namely the users who are going to update the RIS-OSH at specific intervals as well as the functionalities of the system. Each step of the process is characterised by the production of documents: the collected project forms become part of the yearly activity plan and will be published on the Internet in the form of call for proposal once they have passed the departmental and institutional evaluation activities. Generally these documents are considered administrative documentation [15, 16]. However if we analyse their content we can distinguish between information that is an essential part of GL documents from data that are used for the execution of the process from a managerial and administrative point of view. The activity plan aims to document the institute's research activity, and as such it is important especially for policy makers and governmental evaluation committees. From a time perspective the analysis of activity plans can highlight the trends of research topics in the field of safety and health. The call for proposal, on the other hand, belongs to GL official documents.

Figure 5: Production of GL documents and the updating process of their information content

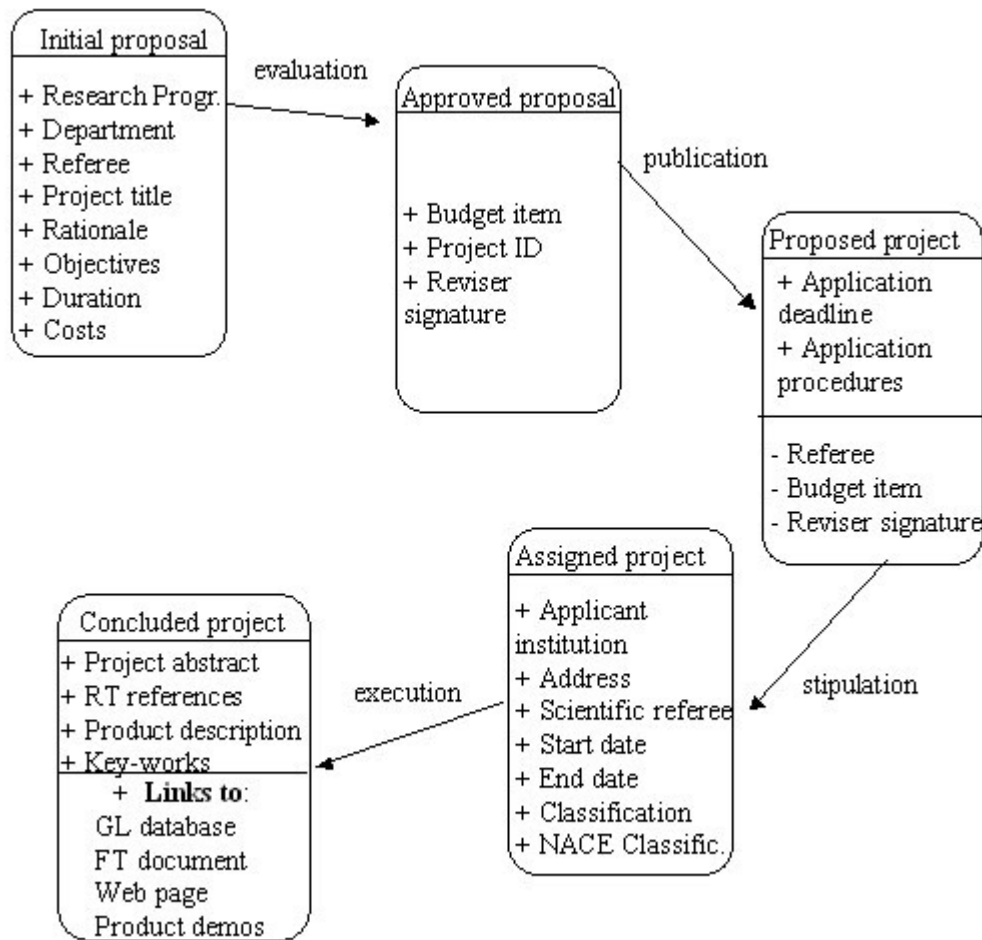


Figure 5 shows the information developed in each step. There is an incremental addition of information in the identified project status described on the top of each information table (*Initial proposal*, *Approved proposal*, *Proposed project*, *Assigned project*, *Concluded project*). The data are added at each step (indicated by a “+” sign in the figure), only at the state of *Proposed project* is there some information that does not appear on the Internet for reasons of privacy (indicated by a “-“ sign in the figure). Data inputs are given at the states of:

- *initial proposal*, when the information is elaborated by the researcher in an Intranet environment of the RIS-OSH system;
- *approved proposal*, when the scientific and administrative evaluations are concluded and the proposal is given an individual identification number (project ID);
- *proposed project*, when the set of proposals are published on the Internet and the applicant research institution can consult the call for proposal and submit its application;
- *assigned project*, when information is inserted at the administrative level and is also available on the Internet;
- *concluded project*, when the ISPESL’s referee evaluates the results and makes the summary of project results, the references of the research report and eventually the description of the research products available on the Internet. The references of the research report will link and update the GL database as well as connect it to the full text of the document, and eventually to available web pages and product demos.

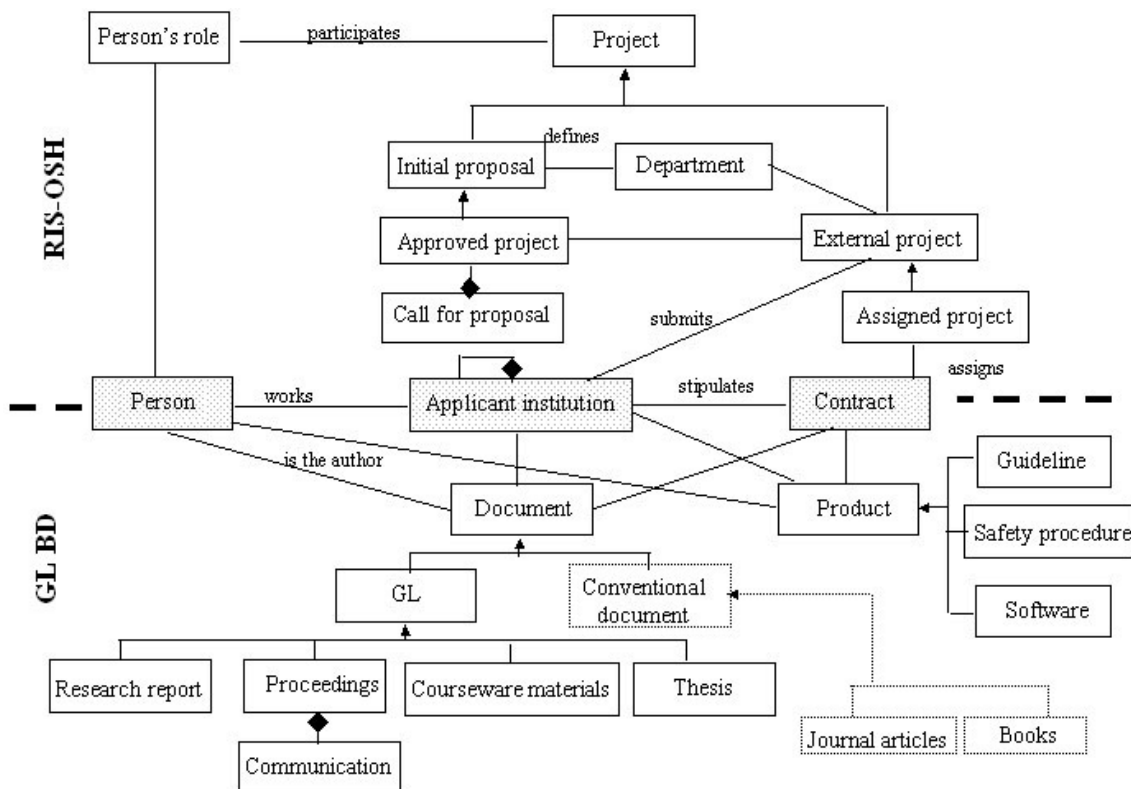
In the RIS-OSH database each status of the project lifecycle is stored together with the related descriptive data showed in fig. 5. This produces different views of information, which

allow reconstruction of the history of the process. In this way the GL documents produced during the process can be retrieved in any time. The consistency of the data are, on the other hand, preserved by the identification of the time of updating, set of data associated to each project status, as well as by the controlled access of the actors, who insert the information.

#### 4. The integration of RIS-OSH with GL database

The integration between the project description and the diffusion of its results is still an open research issue. The solution that we propose is based on the results obtained in a previous study related to the producers and users of GL documents in the field of safety and health [17]. This study has allowed us to identify the bibliographic element, which characterises both research results (deliverables, final research reports) and products (guidelines, courseware materials, technical opinions, safety procedures).

Figure 6: Data Integration of RIS-OSH and GL database



The integration is therefore based on the consistency of the descriptive elements of the two systems (entities, attributes, relations, definition domains and semantic constraints). In fact RIS-OSH and the GL database share the concepts of *person*, *applicant institution*, *contract* (fig. 6). *Person* is specified by the entity *Person's role*, which indicates the role-played within the project: ISPEL's referee, co-ordinator, external scientific referee, contact person, and collaborator. The project is described by different attributes, according to its different states, each status is connected with a specific relation, and for example the call for proposal is composed by approved projects, while assigned project is based on a contract. From the point of view of the GL database both documents and results refer to a *person* (author of the publication and/or product), to the institutions (*applicant institution*), where the project has been carried out (author's affiliation and corporate source) and to a related *contract*. This can also be applied to the entity *product*. There are different types of both documents and products and

each one has its own bibliographic description. Conventional documents can be also produced within the project; the dotted line in figure 5 indicates further links to databases of conventional literature.

## Conclusions

RIS-OSH is currently at developmental. At the moment the modules related to the presentation and evaluation of the proposals have been implemented using the relational database Access and will be tested by the Documentation and Information Department.

The results obtained so far are related to the identification of the GL documents produced during the project lifecycle and their function within the managerial and administrative process as well as documentation of the project. This has been obtained through the analysis of their information content and through the identification of the main steps of the process and of the actors, who are going to update the information on the project.

Moreover we propose a solution to the issue of relating the project description with its results based on the integration of the information contained in RISH-OSH and the GL database. This would provide the user with an information overview, which inserts the results within the context framework of a project, and enable these documents to be connected with others produced in different research projects. This would increase the visibility and the diffusion of the research results and improve their exploitation. Of course the technological transfer is a complex and difficult task, which seeks to initiate the transformation from “information to knowledge” than in turn can lead to “new insights for wealth creation and improvements in the quality of life” [18].

## References

- [1] Jeffery Keith G, Asserson Anne, Revheim Johanne, CRIS, *Grey Literature and the Knowledge Society*, CRIS 2000
- [2] Jeffery Keith G, *An architecture for Grey Literature in a R&D context*, Proc GL99 New Frontiers in Grey Literature, Washington, 4-5 Oct. 1999. Amsterdam, GreyNet, 1999.
- [3] UNISIST, *Guidelines on the Conduct of a National Inventory of Current Research and Development Projects* (Ref. SC/75/WS/13) Paris, March 1975
- [4] <http://www.cordis.lu>
- [5] Zimmerman Eric H, *CRIS-Cross: Current Research Information Systems at a crossroads*, Proc. CRIS 2002 Gaining Insight from Research Information, Aug. 29-31, Kassel. <http://www.uni-kassel.de/CRIS2002/programme.ghk>.
- [6] Lopatenko Andrei, Asserson Anne, Jeffery Keith G, *CERIF, Information retrieval of Research Information in a distribute heterogeneous environment*. Proc. CRIS 2002 Gaining Insight from Research Information, Aug. 29-31, Kassel. <http://www.uni-kassel.de/CRIS2002/programme.ghk>
- \*[7] Adamczak Wolfgang, Bergmann Heinz, Stefani Sven, *Research report online as portal to a wider CRIS*, Proc CRIS 2000, Helsinki, 25-27 May, 2000, <http://www.cordis.lu/cris2000>
- [8] Staab Steffen, Studer Rudi, Sure York, Volz Raphael, *SEAL - a Semantic portAL with content management functionality*, Proc. CRIS 2002 Gaining Insight from Research Information, Aug. 29-31, Kassel. <http://www.uni-kassel.de/CRIS2002/programme.ghk>

- [9] Lepore Benedetto, Cantoni Lorenzo, *Is there any user for this CRIS?* Proc. CRIS 2002 Gaining Insight from Research Information, Aug. 29-31, Kassel. <http://www.uni-kassel.de/CRIS2002/programme.ghk>
- [10] Lepore Benedetto, Cantoni Lorenzo, *op. cit*
- [11] Common Research Information Format CERIF 2000, *CERIF 2000 Guidelines. Final Report of the CERIF Revision Working Group Co-ordinated by DG XIII-D.4, European Commission*, <http://www.cordis.lu/cerif>, Sept. 1999.
- [12] Asserson Anne, Jeffery Keith G., Lopatenko Andrei, *CERIF: Past, present and future: An overview*, Proc. CRIS 2002 Gaining Insight from Research Information, Aug. 29-31, Kassel. <http://www.uni-kassel.de/CRIS2002/programme.ghk>
- [13] Krause Jürgen, *Current Research Information as part of digital libraries and the heterogeneity problem integrated searches in the context of databases with different content analysis*, Proc. CRIS 2002 Gaining Insight from Research Information, Aug. 29-31, Kassel. <http://www.uni-kassel.de/CRIS2002/programme.ghk>
- [14] Foler M. UML distilled. Addison-Wesley, 1997.
- [1ç] Dovigo A., Marzano, G. *La letteratura grigia e dinamiche documentali*, Bollettino d'informazione AIB, 30 (3/4), 1990
- [16] D'Auria Gaetano, *La letteratura grigia tradomanda di trasparenza amministrativa e offerta di conoscenze sull'amministrazione*, I convegno sulla LG, Roma, 4-5 giugno, 1999, ISTISAN Congressi, 29, 1993.
- [17] Aceti Angela, Castriotta Maria, Di Cesare Rosa, Daniela Luzi, *A profile of GL producers in the field of safety and health at workplaces in Italy: results of a sample survey*, Proc GL99 New Frontiers in Grey Literature, Washington, 4-5 Oct. 1999. Amsterdam, GreyNet, 1999.
- [18] Research Councils, *Long term technology review of the science and engineering base. Full report*, April 2000, [www.rcuk.ac.uk/ltr/finalversion/LTTR\\_InformationToKnowledge.htm](http://www.rcuk.ac.uk/ltr/finalversion/LTTR_InformationToKnowledge.htm)

---

## Author Information

### Daniela Luzi

CNR, Consiglio Nazionale delle Ricerche  
Istituto di Ricerche sulla Popolazione e le Politiche Sociali, IRPPS  
Via dei Taurini, 19  
00185 Rome, Italy

[l.libutti@irpps.cnr.it](mailto:l.libutti@irpps.cnr.it)  
[d.luzi@irpps.cnr.it](mailto:d.luzi@irpps.cnr.it)

## Grey Literature in the Life of GESAMP, an International Marine Scientific Advisory Body

Bertrum H. MacDonald<sup>a</sup>, Ruth E. Cordes<sup>a</sup>, and Peter G. Wells<sup>b,c</sup>

<sup>a</sup>School of Library and Information Studies, and <sup>b</sup>School for Resource and Environmental Studies, Faculty of Management, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5

<sup>c</sup>Environmental Conservation Branch, Ecosystem Science Division, Environment Canada, 45 Alderney Drive, Dartmouth, Nova Scotia Canada B2Y 2N6

**Keywords:** Grey Literature; Information Use; Citation Analysis; GESAMP; United Nations; Marine Sciences

**Abstract:** GESAMP, an international marine scientific advisory group sponsored by several UN bodies, has published significant reports on marine pollution and marine environmental protection since it was established in 1969. Although thoroughly reviewed and refereed, many of GESAMP's publications fit within the internationally-accepted definition of grey literature. Since grey literature can be difficult to identify and locate, are GESAMP's publications ever used? GESAMP serves as an exceptionally good organization to test for access to and uses of grey literature. Through an analysis of its publications this paper shows that even when an organization relies on grey literature as its primary means of output, the publications can receive extensive use. Nevertheless, problems of identification and access remain, and are not fully addressed by digital publication. Many of these problems could be minimized by following recommendations included in the paper.

**Acknowledgements:** Support for this study was provided by the GESAMP secretariat, IMO, London; Dalhousie University; and Environment Canada. Elaine Toms, Faculty of Information Studies, University of Toronto, provided helpful comments on the manuscript, and Peter Cordes assisted with some of the data extraction.

### Introduction

For over thirty years, GESAMP, a leading international scientific advisory body on marine pollution and marine environmental protection, has been publishing significant reports. Often printed in limited press runs, the reports are, nonetheless, comprehensive assessments of the health of global marine environments by scientific experts. Produced by an organization that is sponsored by intergovernmental United Nations agencies in "formats not controlled by commercial publishing," these reports clearly fit the internationally-accepted definition of grey literature.<sup>1</sup> Furthermore, the idiosyncratic publishing features of GESAMP reports, coupled with their handling by indexing services, highlight the complexity and problems of this genre. For decades researchers and organizations have grappled with handling publications of this sort (Posnett and Baulkwill, 1982; Auger, 1998; Luzi, 2000; Søndergaard, Andersen, & Hjørland, 2003). Yet, as our analysis of the GESAMP reports shows, problems of identification and access continue even when recent major developments and advancements in digital publishing might resolve some of the issues.

In the same period that GESAMP has been publishing, other "governments, intergovernmental groups (e.g., the United Nations and its agencies, ICES<sup>2</sup>, OECD<sup>3</sup>) and non-governmental groups around the world have reported...on the overall condition or state of the marine environment" (Wells, 2003, p. 1219), one of the key functions of GESAMP. Therefore, why should any interest be paid to the publication and diffusion of GESAMP reports on similar topics? Beyond the usefulness of the reports to policy and decision

---

<sup>1</sup> Grey literature was defined by the Third International Conference on Grey Literature (1997, p. iii) as "that which is produced by all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers."

<sup>2</sup> International Council for the Exploration of the Sea

<sup>3</sup> Organisation for Economic Co-operation and Development



makers in the UN agencies, is anyone in the wider marine studies and policy communities, or other researchers, using the reports? If the reports were ignored, would it matter? These questions became important to GESAMP itself as over the past three years the organization undertook an evaluation of its global effectiveness and future.

Grey literature is central to the life of GESAMP. Technical reports, prepared via thorough refereeing and review by a large number of scientific experts, but not published by commercial publishers, are the primary means by which GESAMP provides advice to its parent agencies and other interested parties. GESAMP's technical reports contain cutting edge reviews and analyses coupled with recommendations about problems threatening the health of marine environments. As a result, these reports significantly augment knowledge on these subjects. For example, GESAMP contributed a research paper (GESAMP, 1991a) to the preparation of Agenda 21, adopted at the United Nations on Environment and Development (UNCED) (Earth Summit '92), held in Rio de Janeiro. GESAMP's hazard evaluation procedures and guidelines contribute to the application of the International Convention for the Prevention of Pollution from Ships, or MARPOL (1973). Modified in 1978, MARPOL 73/78 continues as a major international convention for regulating ocean pollution by oil and chemicals (Wells, Höfer, & Nauke, 1999). GESAMP's most recent publication is a revision of these guidelines (GESAMP, 2002).

Since there are no firm publishing rules for grey literature, GESAMP was not constrained in its dependence on this genre. Further, although the cost of preparing, producing, and distributing the technical reports is high (estimates of over US\$1m for some reports have been suggested), no effort was taken previously to determine whether this form of report preparation and distribution was efficient or effective. Judging a need of GESAMP, the authors engaged in a thorough study of its publishing history and an extensive analysis of the use of its reports. Our recommendations should help GESAMP improve its publication practices, and have a positive impact on awareness and use of its publications. As an interdisciplinary research team undertaking this study, we have found that our questions and approaches are evolving as we consider findings. Ultimately, the study aims to track the influence of the GESAMP reports; however, determining actual influence is very difficult to confirm systematically. As a significant step in this study, this paper reports on an extensive analysis of the production and use of publications by GESAMP (further details are found in Cordes, 2003), an organization that can act as an exemplar for discussion of publishing implications where research is grey dependent.

### History and Operation of GESAMP

The Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP) has fulfilled an important international scientific role since it first met in 1969 at the Intergovernmental Maritime Consultative Organization (IMCO) headquarters in London (Pravdić, 1981; Windom, 1991; Wells, Duce & Huber, 2002). The Group consists of natural and social scientists sponsored by eight UN bodies.<sup>4</sup> The GESAMP "mechanism encourages collaboration and coordination of activities within the UN system on matters relating to marine environmental protection" (Wells, Duce, & Huber, 2002, p. 79).

GESAMP's original mandate, which focussed on scientific aspects of marine pollution, has expanded to include periodic assessments of the state of the marine environment, and, since 1993, matters concerning the protection and management of marine living resources and ecosystems, especially where the multidisciplinary advice proves beneficial (Wells, Duce, & Huber, 2002). The GESAMP secretariat, made up of representatives of the sponsors, appoints members to provide advice on marine environmental issues of concern to them and to maintain an appropriate disciplinary and geographical balance. Members (to date, maximum of 32—four per agency) act independently in their personal capacities as marine specialists. In the first twenty years of its history, GESAMP had 103 members from 36 countries (Windom, 1991, p. 7). All members are recognized researchers in their fields, chosen on the basis of their reputation and their ability to hold a broad perspective. The broad multidisciplinary makeup of GESAMP has been one of its significant strengths.

At GESAMP's annual sessions, members review their work program, receive and, after thorough

---

<sup>4</sup> GESAMP is sponsored by the International Maritime Organization (IMO) (formerly IMCO); the Food and Agriculture Organization of the United Nations (FAO); the United Nations Educational, Scientific and Cultural Organization (UNESCO), and its Intergovernmental Oceanographic Commission (IOC); the World Meteorological Organization (WMO); the World Health Organization (WHO); the International Atomic Energy Agency (IAEA); the United Nations (UN); and the United Nations Environment Programme (UNEP). The number of sponsors grew from four in 1969 to eight in 1977, but WHO and UNEP currently limit their involvement.

consideration, approve publication of reports, and discuss emerging issues affecting the oceans. Most of the substantive work on specific issues, however, is carried out inter-sessionally by designated working groups under agency sponsorship. Each group is chaired by a GESAMP member, and includes invited marine specialists from around the world. For example, “during the 1999-2000 inter-sessional period, 39 experts from 22 countries participated in GESAMP working groups” (Wells, Duce, & Huber, 2002, p. 79). Meetings of the working groups are highly technical, with the goal of producing reports on very specific topics. Most groups are ad hoc, with specific tasks that can be accomplished in one to three years, and disband after their reports are reviewed, revised and published in the GESAMP *Reports and Studies* series. Some groups have had lengthy histories, however, and have produced many reports. For example, the Working Group on the Evaluation of the Hazards of Harmful Substances Carried by Ships (known as the EHS Working Group), began as an *ad hoc* panel in 1972, but since 1974 has had a major role evaluating the hazards of chemicals carried by ships for the International Maritime Organization and the MARPOL Convention (Wells, Höfer, & Nauke, 1999).

GESAMP has issued over 70 reports, including major periodic assessments, such as *The State of the Marine Environment* (# 39) (GESAMP, 1990). One, *A Sea of Troubles* (# 70) (GESAMP, 2001), was designed to reach a decision-making and public audience, and a science journalist was hired to draft and finalize the text, with extensive GESAMP input and external review. Many scientists think of grey literature or technical reports as work that has not been peer reviewed, but that is emphatically not the case for GESAMP’s reports. Each report is drafted in one of the working groups; report # 64 (GESAMP, 2002), for example, was the result of six years of work by the thirteen member EHS Working Group. Frequently, the chair of a working group arranges for a late draft of the report to be refereed by external reviewers; a draft of report # 64 was refereed by nine external experts. The working group submits a final draft to an annual session of GESAMP, in print and in an oral presentation, where further detailed, page-by-page consideration is undertaken. Draft reports are often returned for modification, and sometimes for major revision. Most early reports did not explicitly acknowledge the thorough, open reviewing process, but reviewing details have been included recently; report # 70, for example, lists over 90 individuals who had various roles in its preparation.

After three decades of work, GESAMP is in a transition period. It has felt directly the funding cuts in the UN system. In addition, increasing pressures on marine environments (GESAMP, 2001; Rozwadowski, 2002), have prompted many organizations to pursue their own lines of research and policy development. Moreover, it is harder to recruit volunteer experts, since employers (governments, especially) are less willing to let their employees take on expensive membership responsibilities. In 2000, the sponsoring agencies “agreed to carry out an independent and in-depth evaluation of the achievements of GESAMP, its impact, scope, membership, working methods and role” (United Nations, 2001). GESAMP’s productivity has been affected as the evaluation of the Group was completed and its positive advice and support considered. Nonetheless, a totally independent global body that can provide unbiased advice about marine environmental issues remains a necessity, and it is likely that GESAMP will continue to evolve rather than be disbanded.

### Identifying GESAMP Publications

Beginning with the *Report of the First Session* (GESAMP, 1969), GESAMP has published a variety of reports and documents. The range of publication types is shown in Table 1, and the topics are reflected in the titles of the most highly cited reports (Table 2). The main output of the Group is its *Reports and Studies* series (GESAMP, 2003), begun in 1975, which includes reports of the annual sessions, and 48 thematic reports prepared by the working groups. Each report is published by the agency that hosted the session or was the primary sponsor of the working group. Twelve thematic reports and three reports of sessions are available electronically through GESAMP’s website (GESAMP, 2003). In addition, GESAMP has released other publications, identified by document numbers, including the reports of the first six annual sessions, numerous report drafts, and other working documents. A report prepared for Earth Summit ‘92 was published as an *UNCED Research Paper* (GESAMP, 1991a). All reports were published in English, and some, including reports of the sessions from 1970 to 1997 and a few of the thematic reports, were translated into French, Spanish, and/or Russian. Recent session reports contain summaries in these languages. One recent thematic report appeared in Chinese. The number of copies in a print run varies widely. Some reports have been published in low numbers (500-600), insufficient to meet distribution requirements (M. Huber, personal communication, November 6, 2002). Others have been produced in much larger numbers. The *Impact of Oil and Related Chemicals and Wastes on the Marine Environment* (GESAMP, 1993), for example, has been reprinted twice (IMO, personal communication), and *A Sea of Troubles* (report # 70) (GESAMP,

2001) had a completely distributed press run of about 6,000. The latter report is now available only in electronic format.

Many reports have appeared in multiple versions. Starting in 1990, seven thematic reports, and one paper from the *Technical Annexes to the Report on the State of the Marine Environment*, were recast and published as monographs or as papers in the primary research journals. For example, in 1997 report # 59 (with additional supplementary chapters) was published as a book by Cambridge University Press (Liss & Duce, 1997) and report #s 45 and 38 were republished as papers in *Marine Pollution Bulletin* and *Global Biogeochemical Cycles* respectively (Gray et al. 1991; Duce et al., 1991). GESAMP reports have also been republished by the sponsoring agencies in other series; identifying these has been challenging, and some may have been missed. One thematic report reappeared in IAEA's *Technical Reports* series. Thirteen more were republished in the *Regional Seas Reports and Studies* series (UNEP, 2003), with one in both Spanish and English. The *Technical Annexes to the Report on the State of the Marine Environment* were published only in that series. Reports of four early sessions were published in *FAO Fisheries Reports*, one in English and three others in Spanish. Seven of the ten most widely cited GESAMP publications (Table 2) were republished in a related book or paper, and/or *Regional Seas* version.

Confirming the identity of the output of GESAMP is made all the more challenging since the publishers of GESAMP reports have placed a low priority on abiding by consistent bibliographic standards. Two reports, #s 22 and 52, lack publication dates, and ISSNs and ISBNs have been irregularly applied. The fact that there are eight different publishing agencies adds to the degree of complexity of identity, access, and citation. Two reports were published by different agencies as # 11, although subsequent GESAMP lists give one the number 12. The titles of successive reports on the same topic are often similar (e.g., *Review of Potentially Harmful Substances...* as in report #s 22, 28, 29, 34, and 42), or identical (e.g., *Interchange of Pollutants Between the Atmosphere and the Oceans*, as in report #s 13 and 23).

Seemingly modest variations in the name of the series and the organization can cause identification problems. Examples of inconsistencies are given in Table 3. The name of the series, *Reports and Studies*, appears (if at all) in that form on the title pages of all reports, and on the covers of reports up to 1990. On the redesigned cover template, used since 1991 for most reports, the title appears as *GESAMP Reports and Studies*. However, the recommended citation format always is *Rep. Stud. GESAMP*. The GESAMP acronym is inconsistently placed before the full name, except in the recommended citation format. Although the acronym has remained unchanged, the official name of GESAMP has changed six times, as sponsoring agencies were added and its mandate expanded, from *Joint IMCO/FAO/UNESCO/WMO Group of Experts on the Scientific Aspects of Marine Pollution* in 1969 to *IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection* since 1993. Since all of these variations are reflected in catalogues and indexes, identifying GESAMP's reports comprehensively and exactly can be challenging.

It can be very difficult for a reader or researcher to be sure that all GESAMP publications on a subject have been consulted. Although publication lists are printed in some of GESAMP's reports, and one is given on the organization's website (GESAMP 2003), inaccuracies occur. Working titles of draft reports may be included, but not be updated when the final report is published. This is the case for the title of report # 64 (GESAMP, 2002) listed on the website (GESAMP, 2003). None of the indexing services (currently, *Aquatic Sciences and Fisheries Abstracts (ASFA)*, *Environmental Science & Pollution Management*, *GeoRef*, and *Zoological Record*) contain records for all of GESAMP's publications. Given its broad mandate, *ASFA* should index all of GESAMP's reports, but seven thematic reports, including two major 2001 reports (#s 70 and 71), and two reports of sessions are missing. In addition, the indexes often introduce their own variations of the corporate author name and series title, complicating searching.

GESAMP reports have been acquired by libraries but distribution problems can affect the completeness of collections. A search of holdings of selected major marine sciences and national libraries, along with e-mail correspondence with librarians at some institutions, confirms that holdings are often incomplete (Table 4). The National Marine Biological Library (NMBL) in Plymouth, England "contains one of the world's major collections of literature on aquatic sciences and fisheries, and acts as a marine biological sciences resource for the UK." This library aims "to give a lead, both nationally and internationally, in the development and provision of information to the wider aquatic community" (National Marine Biological Library, 2003). The library is missing only two GESAMP publications, from the 1980s. Marine sciences libraries, and the British Library Document Supply Centre, have more complete collections than the Bibliothèque nationale de France and the Library of Congress in the United States. The Library of Congress holdings of GESAMP reports since 1990 focuses on aquaculture and other coastal issues. Dalhousie University Libraries, which are a UN depository, did not receive copies of report # 70 (GESAMP, 2001) and

its companion # 71, of which insufficient copies were printed (M. Huber, personal communication, November 6, 2002). Spotty holdings of libraries pose both awareness and access difficulties for researchers at those institutions. If these libraries are an indicator, then it is possible that lack of awareness of some GESAMP publications could be widespread. This could be a major problem in the world of grey literature and is worth further investigation.

### Tracking the Use of GESAMP Publications

Determining who uses GESAMP publications is not a trivial task. While “use” can be defined in a number of ways, the most systematic, albeit incomplete, method of tracking use is to locate citations to the GESAMP reports and other publications. Citation indexes are available for much of the journal literature, but no comprehensive tool exists for citations found in either grey literature (print or electronic) or in monographs.

Tracking citations found in journal literature can be achieved via Web of Science, a citation index, compiled by the Institute for Scientific Information (ISI) in Philadelphia, containing citation records from more than 8,000 periodicals in the sciences, social sciences, arts and humanities. This index was searched for citations to the English language publications outlined in Table 1, but locating citations to grey literature is not easy. In Web of Science’s compact citation format, 18 characters are allowed for the cited author, and 20 for the cited work. The cited author field may be blank, or may contain a variation of the name of GESAMP or a sponsoring agency, e.g., “GESAMP,” “JOINT GROUP EXP S,” “GROUP EXP SCI ASP,” “IMO FAO UNESCO WM,” or “UNEP.” The cited work field may begin with a report number, or may contain a report title, a variation of GESAMP’s name, or the report series title, e.g., “115 REG SEAS REP STU,” “STAT MAR ENV,” “IMOFAOUNESCOWMOWHIOIA,” “REP STUD,” or “GESAMP REPORTS STUDI.” Citations to document GESAMP XVII/5 were variously coded as “GESAMPXVII5 UNESCO” and “GESAMP175 UNESCO FIN”. (Documents are given identification numbers based on the numbers of the session and agenda item where they are discussed; GESAMP XVII/5, the final draft of report # 32, was the fifth agenda item at the seventeenth session in 1987.)

Authors commonly introduce or perpetuate errors in citation data, which are not detectable until citations are compared to original publications. While checking inconsistent citations, papers were found to contain citations with errors in report numbers, titles, and publication years. In ambiguous cases, the report title was assumed to be the most accurate information. Web of Science also introduces occasional coding errors, including misspelling GESAMP (GEASMP, GSAMP), reversing digits (report # 59 became # 95), truncating data (report # 43 became # 4), putting data in the wrong field (report # 38 in 1989 became a report in 1938, and “Duce RA” became a “cited work” entry), and combining similar adjacent citations (two citations to GESAMP reports # 21 (1984) and # 31 (1987) appear as one citation to report # 31, 1984).

Retrieval of relevant citations in Web of Science improves as a searcher discovers more variations in how authors cite the GESAMP publications and how the citations are entered in Web of Science. Errors like those noted above emphasize human fallibility often found in citation data, and underscore the need to check problematic citation data against the original citing paper and GESAMP’s publication list. Creativity and sheer persistence are needed to locate grey literature citations in Web of Science!

### Uses of GESAMP Publications

Although GESAMP publications may be difficult to find in libraries and indexes, citation data confirms that the GESAMP publications are definitely used. A comprehensive search for citations using Web of Science (for details on the methodology, see Cordes, 2003) located 1518 citations to GESAMP publications in 1251 papers indexed by July 28, 2002. Since Cordes (2003) was written, the citation data has been updated to include one target publication (GESAMP, 1991a), a citation to it, and 81 other citations. Over 88% of the citing papers contain only one citation to a GESAMP publication, but three papers about GESAMP each have between 15 and 17 citations. Figure 1 shows the number of citations per year. Before 1991, only one year (1982) had more than 20 citations. Between 1989 and 1991, 13 thematic reports, a book and three papers were published, which led to an accelerated citation rate through the 1990s, to 170 citations in 1999. Although a search for citations in 2002 publications, indexed after July 28, has not been thorough, citations to GESAMP publications in that year are even higher. Most of the citations (96%) make reference to versions of the thematic reports, rather than to reports of the sessions. This statistic provides evidence that the thoroughly prepared reports are the most important GESAMP publications in the eyes of readers. The ten most cited reports in their various versions (Table 2) account for 1029 citations, two thirds of the total. The eight books and journal articles related to the thematic reports received 38% of the citations,

showing a decided bias toward these items; one paper (Duce et al., 1991) received 324 citations, 21% of the total. Overall, citations were found to 85 of 115 identified English language GESAMP publications, and to numerous working documents and drafts of reports. No citations were found to two of the thematic reports (#s 20 and 36), a point which remains unexplored.

The strength of citations over time for the two most highly cited GESAMP reports (Table 2) is illustrated in Figures 2 and 3. *The Atmospheric Input of Trace Species to the World Ocean* received an increasing number of citations for more than a decade after its publication in 1989, with the highest number of citations in a single year occurring ten years after its release. Citations are even higher in 2002, when citations from publications indexed after the initial data collection date of July 28, 2002 are included. The journal version of this report (Duce et al., 1991) received 79% of the 410 citations. The 22 co-authors (members of the working group) led this trend, with 57 citations to the journal version among their 67 self-citations. Citations to *The State of the Marine Environment* (Figure 3) show a bimodal citation pattern, but it too received more citations a decade after it was published than in the first two years following its release. The UNEP *Regional Seas* version of this report, preferred by its working group members for self-citation, received more citations than the book version. These examples demonstrate that GESAMP publications continue to hold the attention of researchers for at least ten years, in contrast to many scientific papers, which if they receive any citations at all, often do so for only a short period after publication. Citation by authors familiar with the report also may be an important factor in promoting awareness, i.e., citations by these authors may influence others to use and cite the reports.

To test whether direct involvement with GESAMP would skew the citation patterns, 690 individuals connected to the Group (GESAMP members and administrators, working group members, reviewers and observers at sessions) were identified. These individuals account for 28.3% of the citing papers and 33.8% of the citations. Therefore, authors outside the organization wrote over 70% of the citing papers.

The geographic distribution of citing authors further characterizes the use of GESAMP publications. An analysis of the countries named in the Web of Science "availability" field for 91% of the citing papers shows the following distribution: Western Europe (43%); North America & Caribbean (31%); Asia (8%); Australia & New Zealand (6%); Eastern Europe (5%); Middle East (4%); Africa (2%); and South America (1%). Thus, most of the researchers citing GESAMP publications are located in Western Europe and North America. The majority of the addresses (53%) are for universities, and most of the rest are for government facilities. This distribution may be biased, however, by the selection of journals, published largely in English in North America and Western Europe, that are indexed in Web of Science.

GESAMP's publications were cited in 311 different journals, with *Marine Pollution Bulletin* containing the most citations (12%). From the categories established by ISI's *Journal Citation Reports*, it can be shown that journals in 84 subject areas cite GESAMP publications, with 56 each containing at least two citations. The top five categories are *Environmental Sciences; Marine & Freshwater Biology; Oceanography; Geosciences, Interdisciplinary; and Meteorology & Atmospheric Sciences*. Farther down the list, but still in the top 25, are the social science categories *International Relations; Environmental Studies; and Law*, and diverse science categories including *Engineering, Environmental; Fisheries; Geochemistry & Geophysics; Public, Environmental & Occupational Health; and Food Science & Technology*. The audience for GESAMP publications is indeed wide.

Although GESAMP reports are cited in monographs (Wells, 2001), the only way to locate such citations is to examine the books themselves. A suggested methodology could include: 1) establish a set of descriptors that objectively identify subject categories in the field of marine environmental protection and related topics, ensuring coverage of scientific, managerial, and popular literature. The set of key words found in the *Marine Pollution Bulletin*, a key journal that cites GESAMP reports, and/or the *ASFA Thesaurus*, could be used as a reasonably objective source of subject descriptors; 2) using these descriptors, search the collections of selected libraries (holdings of marine science, oceanography, management, public policy, etc.) for relevant monographs; 3) create a list of the monographs (possibly establishing sample subsets); and 4) examine references and citations noting all those referring to GESAMP publications. This method would be comprehensive, but very time-consuming. A preliminary survey (Wells, 2001) has demonstrated that established authors and major publishing houses do refer to GESAMP reports; 39 citations to 19 of GESAMP's thematic reports, and two citations of its definition of marine pollution, were found in 22 relevant books published from 1987 to 2000.

### **Publishing Recommendations for GESAMP**

As this paper shows, an array of grey literature issues was found while examining the production and use of GESAMP reports and other publications. The following recommendations, which are designed to reduce

description inconsistencies, and to improve awareness, retrieval, and ultimately use, of the GESAMP reports, have been forwarded to the GESAMP secretariat:

- Emphasize the acronym GESAMP by placing it at the beginning of the full name of the organization. The acronym by itself could become the official name, like some other UN agency names, e.g., UNESCO, UNICEF.
- Use only the acronym in the recommended citation format, e.g., GESAMP. (2001). *A Sea of Troubles. Rep. Stud. No. 70.* (35 pp.).
- Establish a standardized name for the series with the acronym GESAMP at the beginning, and apply it consistently on cover titles and title pages, and in the recommended citation format.
- Ensure that the title of each report is consistently the same in all locations where it is found: cover of a report, title page, recommended citation format, publication lists within printed reports, and on the GESAMP website.
- Obtain a new ISSN when a name change occurs in GESAMP or in the report series.
- Always include an ISBN in each published report, along with the ISSN.
- Distribute new reports widely, with generous publicity.
- Ensure that a copy of each new report is sent to each indexing agency.
- Alert ASFA to the reports not already indexed.
- Describe the peer review process in each report, and on the GESAMP website.
- Consider publishing the reports in two series, one for thematic reports and one for reports of sessions, since the production and reviewing processes for the two types are markedly different.
- Give much greater effort to ensure that the reports, especially the thematic ones, are translated, since GESAMP is a global advisory body sponsored by the UN.
- Prepare book or journal versions of new thematic reports routinely.
- Keep the website up-to-date, accurate, and relevant to potential users, to encourage use.
- Continue creating electronic versions of older reports, especially thematic ones.
- Centralize the production and distribution of the reports by a single agency or the GESAMP secretariat, if possible, to make it easier to carry out the other recommendations.

### Conclusions

Many intergovernmental organizations and non-governmental scientific bodies publish and depend upon grey literature. Whether such literature is detrimental to the dissemination of scientific research and advice, and consequently, the impact of the organization, is often not a major concern of such agencies. However, since grey literature by its nature can remain hidden from its potential users, reliance on grey literature as the major means of publication should concern IGOs and NGOs. Even when publications are now made available digitally, as GESAMP has provided in recent years (GESAMP, 2003), the literature can still remain hidden from potential users.

GESAMP has proven to be an exceptionally good organization to test for access to and uses of grey literature. As an international organization, sponsored by the UN and several of its agencies, it has published most of its output as grey literature for over thirty years. Its publishing practices demonstrate how complex and idiosyncratic a report series, with even a small number of reports, can be. While standardization of publishing practices could be the responsibility of GESAMP's secretariat, this paper has shown that human fallibility occurs in citation and indexing well outside the control of GESAMP. Both GESAMP's own practices and external factors impact on accessibility of its reports.

Recent studies (Toms, 1994; Rozwadowski, 2002) have pointed to a growing production of grey literature by major government departments and international intergovernmental organizations. Some countries, e.g., the USA, illustrated by NTIS (National Technical Information Service), depend heavily upon grey literature. If these studies indicate a trend, then increasing reliance on grey literature will become more common in such bodies. The advent of new dissemination models, such as the open access movement emerging in the sciences (Malakoff, 2003), may encourage increased production of grey literature since this genre will be seen as fitting within the open access paradigm. Moving documents to an open access digital environment, however, does not reduce the need for attention to the recommendations supplied to GESAMP by the authors for its publications in paper format. Accessibility does not automatically improve because a publication is placed on an open website; once found large files can be difficult to download by users with limited bandwidth. A search for links to the GESAMP website confirms that there are very few links to the digital versions of the reports (E. Toms, personal communication, November 22, 2003). For example, a search in Google (<http://gesamp.imo.org/link>) retrieves only 27 links, and only one link to the publications section of the GESAMP site (<http://gesamp.imo.org/publica.htm/link>), which is to a GESAMP

publication itself. A further search in Google for related sites (<http://gesamp.imo.org/related>) retrieves only 52 links, some of which are internal links to GESAMP publications.

As this paper has demonstrated, GESAMP's output has been used quite widely. The distribution of citations highlights the lasting importance of these reports, showing that they, and publications based on them, have been referenced in scientific and policy periodicals and books. Some GESAMP reports received sustained citation for more than a decade, in an extensive range of disciplines, by individuals working around the world, most of whom have no direct connection to GESAMP. Grey literature is an important medium, but versions of reports published in the primary journal papers and books often receive more attention. GESAMP studies have advised influential global environmental conferences from Stockholm in 1972 to Rio de Janeiro in 1992 and Johannesburg in 2002. Although use of GESAMP publications can be shown, the larger issue of influence of the reports is more difficult to ascertain. Determining the impact of the reports requires different methods of investigation, including tracing how the reports were cited, interviews with policy makers world-wide, and surveys of researchers. As an examination of the role of GESAMP as a multifaceted global advisory body moves forward, an analysis of influence, particularly in the policy community, will be the focus of attention. Understanding GESAMP's grey literature output and citations to it is central to this new task.

## References

- Auger, C.P. (1998). *Information sources in grey literature* (4<sup>th</sup> ed.). London: Bowker Saur.
- Cordes, R. (2003). Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an international marine scientific advisory body. *Canadian Journal of Information and Library Science* 27(3): 109-127. See also an erratum in 27(4).
- Duce, R.A., P.S. Liss, J.T. Merrill, E.L. Atlas, P. Buat-Menard, B.B. Hicks, J.M. Miller, J.M. Prospero, R. Arimoto, T.M. Church, W. Ellis, J.N. Galloway, L. Hansen, T.D. Jickells, A.H. Knap, K.H. Reinhard, B. Schneider, A. Soudine, J.J. Tokos, S. Tsunogai, R. Wollast, & M. Zhou. (1991). The atmospheric input of trace species to the world ocean. *Global Biogeochemical Cycles* 5(3): 193-259.
- Fowler, S.W. (1990). Critical review of selected heavy metal and chlorinate hydrocarbon concentrations in the marine environment. *Marine Environmental Research* 29(1): 1-64.
- GESAMP. (2003). *GESAMP publications*. Available: <http://gesamp.imo.org/publicat.htm>. Accessed: November 20, 2003
- GESAMP (IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). (2002). *Revised GESAMP hazard evaluation procedure for chemical substances carried by ships*. London: International Maritime Organization. Reports and Studies, no. 64.
- GESAMP (IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). (2001). *A sea of troubles*. Arendal, Norway: GRID-Arendal, United Nations Environmental Programme. Reports and Studies, no. 70.
- GESAMP (IMO/FAO/UNESCO/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). (1993). *Impact of oil and related chemicals and wastes on the marine environment*. London: International Maritime Organization. Reports and Studies, no. 50.
- GESAMP (Group of Experts on the Scientific Aspects of Marine Pollution). (1991a). *Some reflections on scientific research on marine issues*. United Nations Conference on Environment and Development, Research Paper No. 11.
- GESAMP (IMO/FAO/UNESCO/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Pollution). (1991b). *The state of the marine environment*. Oxford: Blackwell Scientific.
- GESAMP (IMO/FAO/UNESCO/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Pollution). (1990). *The state of the marine environment*. Nairobi: United Nations Environment Programme. Reports and Studies, no. 39, and UNEP Regional Seas Reports and Studies, no. 115.
- GESAMP (Joint IMCO/FAO/UNESCO/WMO Group of Experts on the Scientific Aspects of Marine Pollution). (1969). *Report of the first session*. London: Intergovernmental Maritime Consultative Organization. GESAMP I/11.
- Gray, J.S., D. Calamari, R. Duce, J.E. Portmann, P.G. Wells, & H.L. Windom. (1991). Scientifically based strategies for the marine environmental protection and management. *Marine Pollution Bulletin* 22(9): 432-440.

- Liss, P.S. & R.A. Duce. (1997). *The sea surface and global change*. Cambridge: Cambridge University Press, 1997.
- Luzi, D. (2000). Trends and evolution in the development of grey literature: A review. *The International Journal of Grey Literature* 1(3); 106-116.
- Malakoff, D. (2003). Opening the books on open access. *Science* 302: 550-554.
- National Marine Biological Library. (2003). *National Marine Biological Library*. www.mba.ac.uk/nmb/. Accessed: November 20, 2003
- Posnett, N.W. and W.J. Baulkwill. (1982). Working with non-conventional literature. *Journal of Information Science* 5: 121-130.
- Pravdić, V. (1981). *GESAMP, the first dozen years*. Geneva: United Nations Environment Programme.
- Rozwadowski, H.M. (2002). *The sea no knows boundaries: a century of marine science under ICES*. Copenhagen and Seattle: International Council for the Exploration of the Sea and University of Washington Press.
- Søndergaard, T.F., J. Andersen, and B.Hjørland. (2003). Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model. *Journal of Documentation* 59(3): 278-320.
- Third International Conference on Grey Literature. (1997). *Perspectives on the design and transfer of scientific and technical information*. (Conference papers). Amsterdam: Transatlantic D F.
- Toms, E.G. (1994). Aquatic and marine science in Canada at a crossroads: A bibliometric analysis of Canadian government publications from 1978 to 1992. In *Coastal Zone Canada '94, Cooperation in the Coastal Zone: Conference Proceedings*, ed. By P.G. Wells and P.J. Ricketts. Coastal Zone Canada Association / Association Zone Cotière Canada, p. 2384.
- United Nations. 2001. *Joint Group of Experts on Scientific Aspects of Marine Environmental Protection holds thirty-first session*. Available: <http://www.un.org/News/Press/docs/2001/sea1727.doc.htm> Accessed: November 20, 2003.
- UNEP (United Nations Environmental Programme). (2003). *Regional Seas Reports and Studies Publication List*. Available: <http://www.unep.ch/seas/main/rsrs.html> Accessed: November 20, 2003.
- Wells, P.G. (2003). State of the marine environment reports—a need to evaluate their role in marine environmental protection and conservation. *Marine Pollution Bulletin* 46(10): 1219-1223.
- Wells, P.G. (2001). *Citation analysis of GESAMP Reports and Studies, and related publications*. Unpublished manuscript.
- Wells, P.G., R.A. Duce & M.E. Huber. (2002). Caring for the sea – accomplishments, activities and future of the United Nations GESAMP (the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). *Ocean and Coastal Management* 45: 77-89.
- Wells, P.G., T. Höfer, & M. Nauke. (1999). Evaluating the hazards of harmful substances carried by ships: The role of GESAMP and its EHS working group. *Science of the Total Environment* 237/238: 329-350.
- Windom, H.L. (1991). *GESAMP: Two decades of accomplishments*. London: International Maritime Organization.



Table 1—Types of GESAMP Publications\*

Type of Content	Type of Publication	Items in English	Translations Identified
Session Reports	<i>GESAMP Reports and Studies</i>	25	59
	Early sessions - documents	6	15
	Early sessions - other series	1	4
	Related journal articles	7	
Thematic Reports	<i>GESAMP Reports and Studies</i>	48	7
	Supplements to early session reports	2	4
	Identical versions in other series	14	2
	Only published in other series	2	
	Report drafts and other working documents (# not determined)		
	Related books (commercial publishers)	2	
	Related journal articles (primary literature)	6	
About GESAMP	Histories of GESAMP	2	
	Journal articles describing GESAMP (not included as targets in the citation study)	6	

\* Adapted from Cordes (2003), Table1, and reprinted with the permission of the University of Toronto Press.

Table 2 – Ten Most Frequently Cited GESAMP Publications, to 28 July 2002\*

Report Title Versions <sup>1</sup> and Publication Dates	Total Citations
<i>The Atmospheric Input of Trace Species to the World Ocean</i> A: # 38 (1989), B: # 119 (1990), C: Duce et al. (1991)	410
<i>The State of the Marine Environment</i> A: # 39 (1990), B: # 115 (1990) (also in Spanish), C: GESAMP (1991b) <sup>2</sup>	165
<i>Technical Annexes to the Report of the State of the Marine Environment</i> B: # 114 (1990), C: Fowler (1990)	108
<i>The Sea-Surface Microlayer and its Role in Global Change</i> A: #59 (1995), C: Liss & Duce (1997)	76
<i>Land/Sea Boundary Flux of Contaminants: Contributions from Rivers</i> A: # 32 (1987)	60
<i>The Review of the Health of the Oceans</i> A: # 15 (1982), B: # 16 (1982)	47
<i>Impact of Oil on the Marine Environment</i> A: # 6 (1977)	44
<i>Review of Potentially Harmful Substances: Arsenic, Mercury and Selenium</i> <sup>3</sup> A: # 28 (1986), B: # 92 (1988)	43
<i>Impact of Oil and Related Chemicals on the Marine Environment</i> A: # 50 (1993)	42
<i>Global Strategies for Marine Environmental Protection</i> A: # 45 (1991), C: Gray et al. (1991)	42

\* Updated and adapted from Cordes (2003), Table 4; reprinted with the permission of the University of Toronto Press.

<sup>1</sup> Versions: A - GESAMP *Reports and Studies*; B - UNEP *Regional Seas Reports and Studies*; C - journal article or book. Citations to draft versions were also included.

<sup>2</sup> Content of the book is identical of that of the report.

<sup>3</sup> Title of *Regional Seas #92: Arsenic, Mercury and Selenium in the Marine Environment*.

**Table 3—Variations (in Bold) in Bibliographic Elements Among the Title Page, Cover, and Recommended Citation Format of Selected Reports**

Report	Element	Location	Contents
# 61 (1996)	Corporate author	Title page	IMO/FAO/Unesco-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection - <b>GESAMP</b> -
		Cover	IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection ( <b>GESAMP</b> )
		Citation format	<b>GESAMP</b> (IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection)
	Series	Title page	Reports and Studies No. 61
		Cover	<b>GESAMP</b> Reports and Studies No. 61
		Citation format	<u>Rep.Stud.GESAMP</u> , (61):66 p.
	Title	Title page	The contributions of science to <b>integrated coastal</b> management
		Cover	The contributions of science to <b>integrated coastal</b> management
		Citation format	The contributions of science to <b>coastal zone</b> management
# 67 (1999)	Title	Title page	Report of the twenty-ninth session, London, <b>United Kingdom</b> , 23-26 August 1999
		Cover	Report of the twenty-ninth session of <b>GESAMP</b> London, 23-26 August 1999
		Citation format	Report of the Twenty-Ninth Session, London, <b>United Kingdom</b> , 23-26 August 1999

Table 4 – Holdings of GESAMP Publications in Selected Major Marine Science and National Libraries

Library	Holdings of <i>Reports and Studies</i>
National Marine Biological Library, Plymouth, England <a href="http://www.mba.ac.uk/nmbl/">http://www.mba.ac.uk/nmbl/</a>	Almost complete holdings (two missing reports will be acquired at R. Cordes's suggestion)
Woods Hole Oceanographic Institute Library, Woods Hole, Massachusetts, USA <a href="http://www.mblwhoilibrary.org/">http://www.mblwhoilibrary.org/</a>	Missing five reports, published 1986-1991 (missing reports will be acquired at R. Cordes's suggestion)
Scripps Institution of Oceanography Library, San Diego, California, USA <a href="http://scilib.ucsd.edu/sio/">http://scilib.ucsd.edu/sio/</a>	Missing several recent thematic reports, and many session reports
Bedford Institute of Oceanography Library, Halifax, Canada <a href="http://www.mar.dfo-mpo.gc.ca/e/library/bio-e.htm">http://www.mar.dfo-mpo.gc.ca/e/library/bio-e.htm</a>	Complete holdings
British Library, London, and Document Supply Centre, Boston Spa, West Yorkshire, England <a href="http://www.bl.uk">http://www.bl.uk</a>	Complete holdings (Document Supply Centre)
Library of Congress, Washington, DC, USA <a href="http://www.loc.gov">http://www.loc.gov</a>	Many gaps in holdings
Bibliothèque nationale de France, Paris, France <a href="http://www.bnf.fr">http://www.bnf.fr</a>	Many gaps in holdings

Figure 1 – Frequency of Citations to GESAMP Publications, to the End of 2002\*

\* Updated from Cordes (2003), Figure 2, and reprinted with the permission of the University of Toronto Press. The updating included 82 citations from 1991 to July 28, 2002, and 120 additional citations from later in 2002.

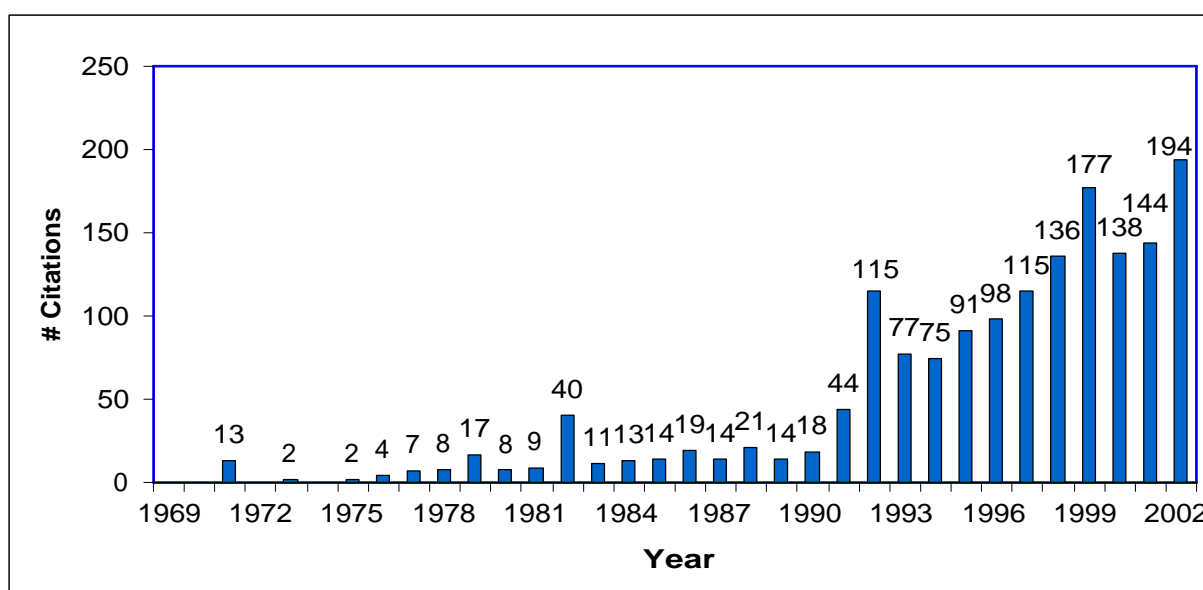
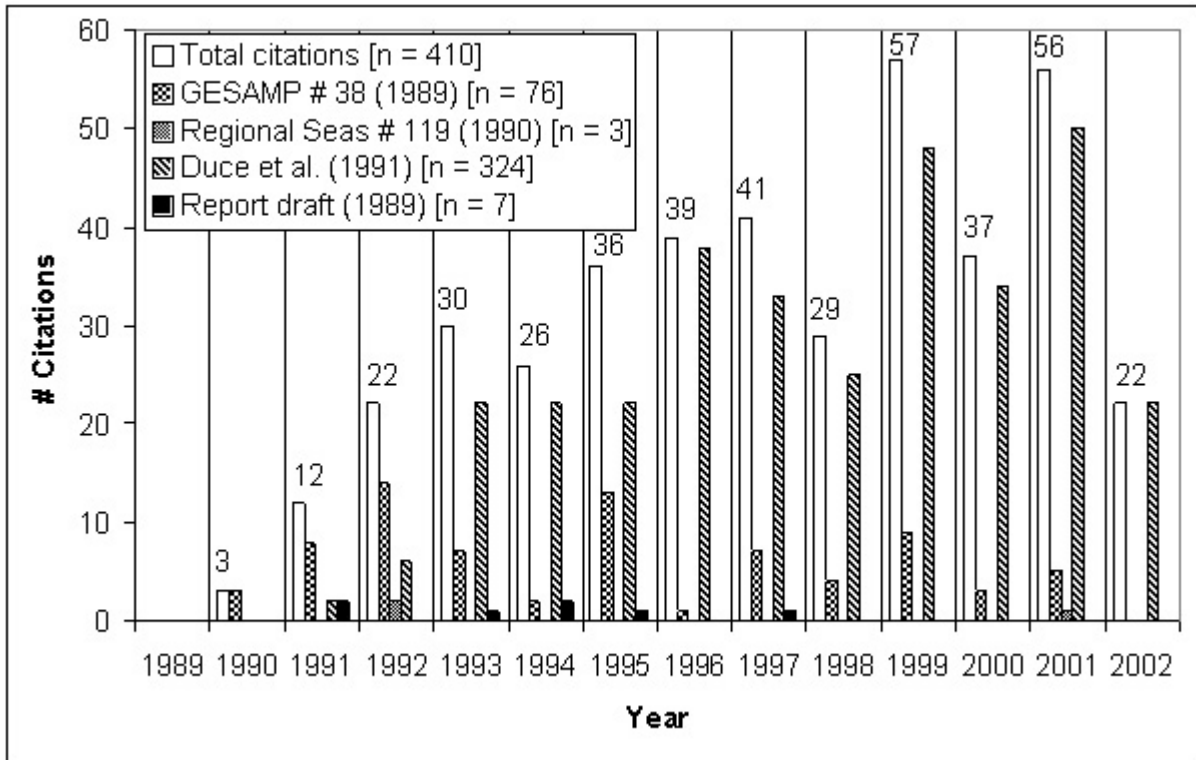


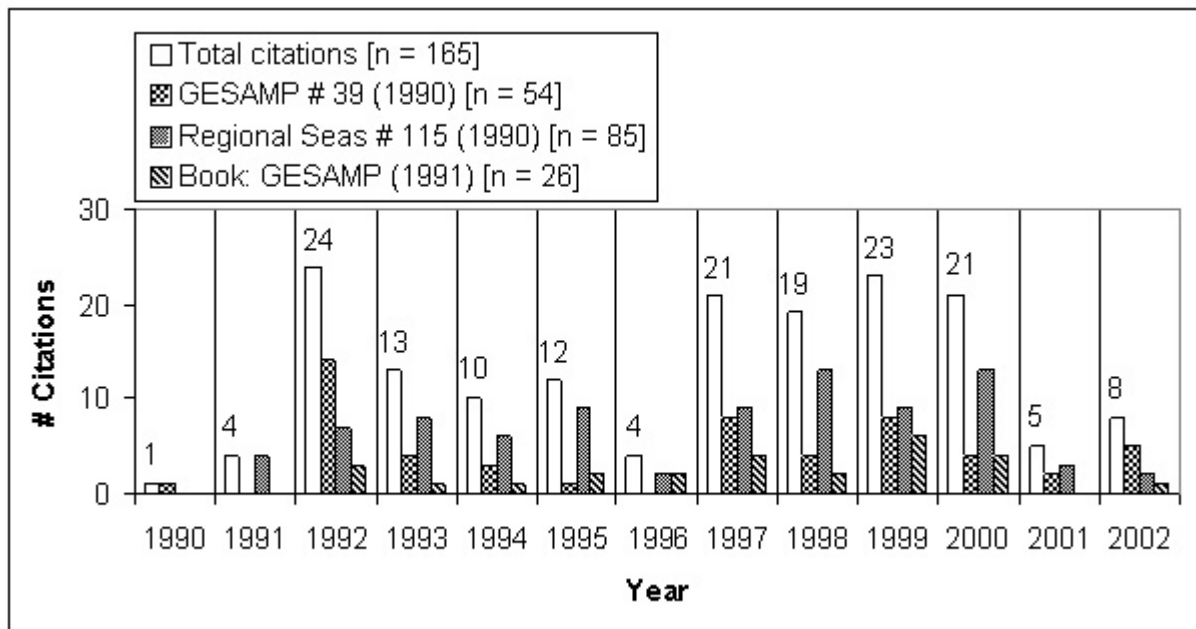
Figure 2—Citations to Versions of *The Atmospheric Input of Trace Species to the World Ocean*, to July 28, 2002\*



\* Reprinted from Cordes (2003), Figure 3, with the permission of the University of Toronto Press. An additional 53 papers, published in 2002, citing Duce et al. (1991) were not included in this analysis.

Figure 3—Citations to Versions of *The State of the Marine Environment*, to July 28, 2002\*

\* Reprinted from Cordes (2003), Figure 4, with the permission of the University of Toronto Press.



**Author Information**

**Dr. Bertrum MacDonald**  
 Faculty of Management  
 Dalhousie University  
 Halifax, Nova Scotia  
 Canada B3H 3J5

bertrum.macdonald@dal.ca  
 www.mgmt.dal.ca

## The Commercialization of Research Findings Documented in Grey Literature

Leonid P. Pavlov

*The Scientific and Technical Information Centre of Russia (VNTIC)*

The problem of economic efficiency of science has traditionally attracted much attention from both the society and the state in Russia. The reason is that the target task of bridging the gap between science and industry has not been achieved and realized. The problem existed and was widely discussed in the Soviet Union and is here to stay in Russia of today. It is quite true that Russia has always been well known for its world level of basic research and scientific achievements but it is evident as well that the industrial implementation of innovations in Russia has always been lagging behind the world standard.

But several other aspects of this permanent problem occurred in a more acute form during the post-Soviet transformation period that has mainly been preoccupied with political unrest however economics and science were also badly damaged by general disorder. The economic efficiency of science emerged on the government agenda again for the following reasons.

First, the budget for scientific research and development decreased dramatically so there was an urgent need for allocating and distributing the “scientific” money most effectively. Second, an overwhelming majority of scientific research and development was (in the Soviet Union) and still is (in Russia) state-funded. This factor is important to be understood when comprehending a different role of the Russian state in intellectual property production and rights distribution as compared to the Western situation. Because of this difference it has not been possible to directly implement the experience and legislation on intellectual property borrowed from abroad, say, from the USA or Europe.

When in the early nineties most of research and development and production organizations and enterprises were almost or completely deprived of the state financial support they were given a relative freedom to be in command of their intellectual products concluding direct agreements to sell or license their technologies, know-how or high-tech products. Several branches of science and industry (e.g. in the areas of aviation and space, nuclear technologies, titanium alloys) have been quite competitive internationally and some of their products were sold abroad directly without even the knowledge of higher state departments to say nothing of observing the state interests both financial and strategic.

So, the problem turned out to be twofold: the bulk of scientific results achieved in the Soviet Union and later on were not commercialized and left uninvolved in real economy (the situation has not been improving - according to the well-informed Moscow newspaper “Industrial Weekly” N 40, November 3 - 9, 2003 due to the existing legal confusion only 0,02% of the total intellectual property volume has been involved in economic operation in Russia now) while some intellectual products, if even a minor part of them, were commercialized but with no participation of the main investor - the state.

In order to improve the situation some important normative documents concerning intellectual property and scientific results were adopted by the legislature and government in the late nineties - early two thousands. To mention but the most significant acts:

- The Decree of President of the Russian Federation N 863 of July 22, 1998 “On the national strategy for involving scientific and technical activity results and intellectual property objects in economic operation”;
- The Enactment of the Government of the Russian Federation N 982 of September 2, 1999 “On the use of scientific and technical activity results”;
- Recommendations of the Parliament Hearings “On the legal support to the intellectual activity” of April 8, 1999 and “The problems of the government protection of intellectual property in the Russian Federation” of May 23, 2000;
- The Conception of the national strategy for involving government-funded scientific and technical activity results in economic operation, published in the leading journal in the field “Intellectual Property” N 7, 2000;

- The Direction of the Government of the Russian Federation N 1607-r of November 30, 2001 “Basic orientations of the national strategy for involving scientific and technical activity results in economic operation”;
- The Enactment of the Government of the Russian Federation N 7 of January 14, 2002 “On the procedures of inventory and cost evaluation of the rights to scientific and technical activity results”.

In accordance with the latest version of the Patent Law passed by the Federal Assembly and signed by President of the Russian Federation in 2003 patent rights can be assigned to the Russian Federation that is to the state as a legal subject. This was impossible under the previous Patent Law of 1992.

In the same period of time, since the early nineties and up to now there has been a heated discussion in the press and within the scientific community over the problems of intellectual property and scientific results rights bearing in mind the low welfare standards of scientists and inventors on the one hand and almost one hundred percent state (government) funding of scientific research on the other hand. Many works on the subject have been published in Russia both by foreign authors (like the classical works by the famous American scientist of the Russian origin, Nobel prize winner Vassily Leontiev on assignment of patent rights on inventions made under government research contracts, first appeared in the mid-sixties) and domestic scientists and men of law many of whom have carefully examined intellectual property rights and patent policies of the USA, European and other developed countries, national and international patent laws.

But still there is no unanimous opinion on the problem and roughly speaking there are two different groups of specialists: those claiming the priority of the state interests and those sticking to the primacy of the personal (author’s) interests. However, such a division is generally typical of the society and does not concern intellectual property rights only. No matter how polar the positions of the experts may be all of them agree to the opinion that the first step should be taken in the direction of complete inventory and registration of intellectual activity results.

While the legal status of research findings documented in patents and publications is well-defined in the Patent and Copyright Laws an immense amount of scientific results contained in grey literature, first of all in scientific reports and dissertations, is practically out of the realm of law or registration. At the same time it is assumed that the grey literature sources should contain a bulk of findings to be commercialized and/or claimed as intellectual property objects.

At present scientific reports and dissertations are being collected in Russia within the framework of the State System for Scientific and Technical Information (abbreviated in Russian as GSNTI). The System includes a hierarchy of information institutions with a number of federal-level centers each responsible for a particular type of scientific and technical information document, e.g. papers from scientific journals, patents, books, serial publications etc. The leading agency of GSNTI for grey literature is the Scientific and Technical Information Centre of Russia (abbreviated in Russian as VNTIC). 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 these also.

The American Technology Preeminence Act (ATPA) of 1991 states that “the head of each Federal executive department or agency shall transfer in a timely manner to the National Technical Information Service unclassified scientific, technical, and engineering information which results from federally funded research and development activities for dissemination to the private sector, academia, State and local governments, and Federal agencies”.

The Federal Law of the Russian Federation “On the obligatory copy of documents” of 1995 (the amended version of 2002) states in Clause 10 on the delivery of an obligatory free copy of unpublished documents (the standard Russian term for grey literature): “The producers of documents shall deliver to the Scientific and Technical Information Centre of Russia (VNTIC) within 30 days an obligatory free copy of:

- scientific reports on research and development projects - since the day of their issuing;
- dissertations - since the day of their presentation and obtaining a scientific degree”.

In Clause 18 the obligations of the organizations that receive an obligatory free copy through a centralized delivery are stated: “The acquisition of the obligatory free copy, its registration and inventory, issuing information publications about it, providing its storage, safety and use are the responsibilities of:

Scientific and Technical Information Centre of Russia (VNTIC) - on the unpublished documents”. (The translation is mine. - L.P.).

The registration of reports and dissertations that is carried out in VNTIC now is rather document- than result-oriented. It would be useful to follow all the lifecycle of a scientific result beginning with the idea and basic research outcome through feasibility study findings to industrial implementation of the result in the form of innovative products and services. In elaboration of the GSNTI a project for the development and design of the National system for scientific results, products and services registration and lifecycle monitoring is being discussed in Russia now. The main idea behind the project is to introduce the scientific results documented in grey literature into the legal space of intellectual property and to monitor the processes of their commercialization and rights transfer.

At this point a few words about the terminology should be said. Unfortunately since the problem of the commercialization of research findings in market economy is relatively new in Russia and of interdisciplinary nature there is no conventional and generally accepted terminology on the subject. Most generally the fruit of science is termed in Russian as “scientific results”. To specify the scope of the term from basic research to applied science it is said about “scientific and technical results”. Bearing in mind the aim of results lifecycle monitoring a generic term “scientific and technical activities result(s)” (STAR) is used to include scientific, scientific and technical results and also products and services which could appear due to scientific research and development activities.

This understanding of STAR combines both the intellectual property objects like inventions with the assigned patents and different kinds of results (scientific, technical, engineering) contained in grey literature, mainly research and development reports and dissertations, describing technologies, methods, devices, designs, etc. and only potentially inventogenic or commercializable.

In 2003 to improve the regulation in the sphere of STAR commercialization the Government of the Russian Federation charged a group of ministries and institutions, VNTIC included, with a mission to investigate the project for the development of National system for unified STAR registration and lifecycle monitoring (further on - the System). The System is intended to be realized on the basis of GSNTI and its information resources with the principal aim of providing the federal executive departments, academic, scientific and business communities with the detailed, reliable and timely information on the effectiveness of federally funded science and the STAR rights and incomes distribution. First of all, this information is necessary to the authorities to make decisions on the national strategies in the areas of science, state budgeting, innovative development and state security.

The System’s macrostructure is determined by the existing departmental and organizational infrastructure. Two subsystems are evident: that of the intellectual property objects (patented inventions, useful models, trade marks, etc.) belonging to the institutions of the Russian agency for patents and trade marks (in Russian - Rospatent) and that of the STAR described in grey literature without the legal patent protection (the competence of VNTIC). In this respect the System is distributed. It is supposed to be designed as a large-scale information system based on network computer technologies and electronic document turnover.

The System is to be centralized at a network architecture level with an all-system information portal, which helps to integrate heterogeneous information resources in a common virtual space. The portal allows for the unified means of navigation, access and STAR registration and updating processes control.



The registration forms used in the current information system and databases of VNTIC contain information on all the scientific reports and dissertations no matter if their content is patentable or unpatentable, useful or useless for commercialization. In 2000 among nearly 6 thousand research and development reports arrived in VNTIC only 266 showed the amount of assigned patents. The distribution of the number of reports with patented results by subject headings is as follows:

Medicine and Healthcare	-	48
Mechanical Engineering	-	29
Chemistry, Chemical Industry	-	25
Agriculture and Forestry	-	24
Biology	-	14
Food Industry	-	10
Electronics and Radio Engineering	-	10, etc.

This information is valuable but it is static (once inputted the forms are never updated), not sufficient for the identification of patentable and/or commercializable STAR as claimed by the authors and not suitable for the STAR lifecycle monitoring that requires registration forms with corresponding supplementable fields. The new unified STAR registration form should include the following information fields grouped in four categories:

1. STAR identifying information
  - unique STAR identification number assigned by the System;
  - source of research funding (if not federal only);
  - actual research and STAR creation cost;
  - kind of research (basic, applied, feasibility, engineering, etc.);
  - federal programme (if any) within which the research is done;
  - document (e.g. federal contract), initiated the research;
  - STAR characteristics - title, abstract, form of presentation (report, dissertation, etc.), organization-the original holder, degree of novelty, lifecycle stage, subject and other classifying codes;
  - patents and other intellectual property protection documents (if any) already assigned in connection with the STAR;
  - document defining STAR use rights distribution and rights holder commitments.
2. Information on research accomplishing organization(s) (STAR authors)
  - head accomplishing organization - name, address and other contact requisites, identifying codes, departmental affiliation, form of property;
  - co-executive organization(s) (if any) - name, address and other contact requisites, identifying codes, departmental affiliation, form of property.
3. Information on government (federal) research funding organization (ministry, state committee, agency, etc.)
  - name, address and other contact requisites, identifying codes.
4. STAR use information
  - used as public information resource (references to publications etc.);
  - used within other STAR (its identification number in the System);
  - used for government (federal) needs in accordance with research contract;
  - used by rights holder in commercial purposes - for production and services output and realization (names, codes and volumes of products and services), for rights transfer or concession to other (juridical) person in accordance with legal agreement (reference to agreement's requisites).

With these forms being filled in by the authors at their computer work stations and entered in the System to replenish the federal STAR register it will be possible to monitor and control the key STAR lifecycle events:

- federal financial allotment to a particular organization to start an ordered research;
- research and development project finishing and documentary report presentation to the financing department and to VNTIC;
- obtained intellectual property objects description, applying for and receiving their legal protection;
- other commercializable STAR identification and description;
- distribution of rights to use STAR between the state, accomplishing organization and/or third persons based on the corresponding agreements;
- rights transfer and licensing to use STAR including for the purposes of production and servicing on their basis;
- observance of the STAR use conditions when the rights holder is the Russian Federation;
- observance of effective STAR use conditions when the rights holder is the accomplishing organization.

Thus, the System is designed to solve the problems of federally funded STAR legal protection and observing the rights and interests balance between the state, the authors, research accomplishing organizations and private investors when involving the STAR into economic operation. However attaining a just balance of interests in intellectual property is an extremely difficult task in principle because of the creative nature of intellectual activity and creative work can hardly be formalized and measured since it has much to do with art, culture, traditions and mentality. Still, it is hoped the System would help to approach better solutions even in the complicated conditions of Russia.

---

## Author Information

---

### **Dr. Leonid P. Pavlov**

Acting Director

The Scientific and Technical Information Centre of Russia (VNTIC)

14 Smolnaya St.

Moscow, 125993

Russia

[pavlov@vntic.org.ru](mailto:pavlov@vntic.org.ru)

[www.vntic.org.ru](http://www.vntic.org.ru)

## Collection development in support of a global information network: A case study of LEISA, Low External Input and Sustainable Agriculture

Wilma Roem, Marilyn Minderhoud

### *Abstract*

*The objective of our presentation will be to show how, over the last 20 years ILEIA has developed a development orientated magazine which currently enjoys an international readership and reputation from primarily grey literature sources. Important elements to be highlighted will be the criteria for collecting and evaluating primary and secondary “grey” source material, the methodology underlying the editorial approach and the role of networking in this process. The context for starting and sustaining the magazine is of course the facilitating environment provided by donor funding but this does not explain the magazines current significance or impact.*

The mission of the Centre for Information on Low External Input and Sustainable Agriculture, ILEIA, is to contribute to the alleviation of poverty, to enhance ecological sustainability and to safeguard social and cultural integrity of smallholders in developing countries through the further development and promotion of LEISA.

*ILEIA has chosen the collection and dissemination of information as a means of contributing to improving the development options available to farmers, by informing those either directly involved in, or indirectly affecting, rural development. ILEIA seeks to be a link that connects local level experiences to global issues and vice-versa, providing a platform for sharing of information and learning from experiences of different countries, regions and continents. The exchange of information has been facilitated through the LEISA Magazine and occasionally through other publications, videos, workshops etc. Information disseminated by ILEIA is freely available to all who wish to use and /or reproduce it.*

*In the 1980s, when ILEIA started its magazine, the initiatives and interesting developments taking place at local level, where only available in the grey sector, in reports, memos and stories. Consistent search for this material throughout the 1980s resulted in a well-known and established collection that had enough status to attract a wide variety of material both grey and from the commercial press publishing development material.*

*After 5 years as a relatively simple informative newsletter, a publication emerged that began more systematically to synthesize and analyze LEISA. Concepts were developed and a network was established. This network and the magazine gradually became a source of information and reference for other writers/researchers. This was in addition to providing instructive and learning material for agricultural practitioners. At the moment LEISA magazine has 20,000 subscribers in 173 countries.*

*Looking back over the years, the LEISA magazine has established a reputation for reliability. Its status today is reflected in the fact that it has generated enough confidence in five organisations in developing countries that they have taken on the task of producing regional editions of the LEISA magazine. Evaluations and request for translation and reprinting rights reflects the status of the LEISA magazine in the world of development orientated publications. Our magazine is also a much used source of material for other more journalistic magazines, web sites and policy documents.*

### **Introduction**

The objective of our presentation will be to show how, over the last 20 years ILEIA has developed a development orientated magazine which currently enjoys an international readership and reputation from primarily grey literature sources. The context for starting and

sustaining the magazine is of course the facilitating environment provided by donor funding but this does not explain the magazines current significance or impact.

The objective of the Centre for Information on Low External Input and Sustainable Agriculture (ILEIA) is to contribute to the alleviation of poverty, to enhance ecological sustainability and to safeguard the social and cultural integrity of smallholder farmers in developing countries by collecting and exchanging experiences on LEISA that can provide solutions to problems of low productivity and environmental degradation by appropriate natural resource management. By collecting, synthesizing and disseminating information on LEISA, ILEIA aims to improve the development options available to farmers by providing a platform for the mobilization and exchange of farm-level experience relevant to those working in smallholder agriculture and in policy and research. ILEIA's documentation and publication activities link local experiences to global issues and vice-versa.

The mission of ILEIA has directed the organization towards open access literature and publishing. Especially for the Third World open access is important. Third World countries often can not afford the high subscription rates of commercial publishers, but worldwide access to information is very important and necessary for these countries. Here knowledge means direct improvement of livelihoods. The increased digital information flow in recent years has facilitated the process of opening up information for a wider public. The Open Access Initiative of 2001 for example is the recent answer of scientific information specialists to the commercial scientific publishers. Open Access opens a worldwide entrance to research data, which has become a necessity for education centers and information centers.

ILEIA's total commitment to an open access approach also influences the way it deals with grey digital information and it strongly supports initiatives such as Open Access 2001 which provides ready access to research data to the global internet-linked research community and seeks to overcome the barriers to access created by the economics of commercial scientific publishing industry.

In the 1980s, when ILEIA started its magazine, information on initiatives to develop technologies that would lower the cost of agricultural inputs and encourage more productive and environmentally friendly farming in poor smallholder communities in the South was difficult to access. Fragmented and distributed in a wide variety of different types of documents it was only available in the grey sector. Mainstream agriculture was not interested in LEISA experiences. ILEIA consistently collected these reports, memos and stories throughout the 1980s gradually building up knowledge of LEISA together with a collection that became well-known in the development community for the quality and range of material it contained. As ILEIA's status increased it attracted an increasing variety of grey literature and gradually, as low input technologies became more widely accepted as a development strategy, books and publications for review from the commercial press were added to the collection.

Being an important resource, the Documentation collection was systematized and its accessibility increased. A comprehensive and uniform system of keywords was developed and introduced and the metadata were stored in a database in 1994. New channels to make the available material on LEISA more widely known and available were developed. Selected bibliographies appeared and later the database was made available on disc.

Important elements in collecting literature are the criteria for collection. What kind of documentation is needed? For ILEIA, documentation is an input that is needed to produce output. The output is formed by: the Magazines, website, publications, book reviews, etc. The output will determine what input is needed. For example, for the Magazine, three types of documentation are needed:

- Documentation as the basis for an article.
- Documentation to further document an article.
- Documentation as reference (further reading) material.

But also the output documents have to be stored in a collection. To make clear what is meant with documentation, two types of documentation have to be distinguished. Depending on the source, documents can be divided in:

- 1 Primary documents: experiences, case studies, best practices, put on paper by people Involved. The produced output documents of ILEIA often belong to this category.
- 2 Secondary documents: research reports, books, conference proceedings, workshop papers. These documents form the input documentation of ILEIA.

Primary documents really are “grey literature”. The objective to produce primary documents on LEISA has led to the establishment of LEISA Magazine as an exchange platform for practical experiences with LEISA techniques. Practical experiences are not easy to be found, they have to be gathered from the field. For this purpose ILEIA developed a strong network with partner organisations to help with that. Two problems have to be solved, first find interesting experiences or cases, second, find someone to write an article about it. For this type of documentation good examples and clear and detailed authors guidelines are needed.

Secondary documents have to be collected from producers all over the world. These documents are sometimes ‘grey’ and sometimes not. For this type of documentation bibliographies, directories and catalogues are needed and recently digital databases and websites. ILEIA has expertise in this and has developed its own acquisition policy.

### LEISA

Material is selected for inclusion in ILEIA’s documentation centre and its magazine and other communication products according to a number of specific criteria. The most important being that the document concerned contains information that can contribute to the alleviation of poverty and the development of smallholder agriculture in the South. Its constituency - its readers, users, and sources in the South - are constrained in access to information by chronic income elasticity that makes the purchase of books, journals, magazines and even newspapers very difficult. The knowledge gap this generates is well recognized by donor agencies and one of the reasons for supporting initiatives such as ILEIA is that it will go some way to alleviate the problem of access to relevant, reliable, unbiased and well-documented information faced by non-governmental organizations, researchers and academics as well as farmers organizations and public government agricultural information agencies in developing countries. Not only is the *LEISA Magazine* free to subscribers in the South but an important element in its information and documentation strategy has been to encourage the free exchange of information and to stimulate the translation and reproduction of the material it publishes in regional languages, local media (community magazines, wall newspapers, posters, training manuals, newspapers etc).

### Documentation for LEISA

In the Collection development of ILEIA three phases can be distinguished. The initiation phase, between 1984 and 1991 starts with collecting secondary documentation. In the 1980s the initiatives and interesting developments on LEISA taking place at local level, were only available in the grey sector, in reports, memos and stories. Criteria for selecting and gathering material for the collection were broad and vague in the beginning. There was not much material available so everything was accepted. A large number of documents in the Dutch language came in during this period. One of the first criteria that narrowed was the language. English became the favoured language, though some interesting documents in other languages have been collected and still are collected. During this period the ILEIA newsletter started as a simple informative newsletter. The LEISA concept was developed and criteria for selection of material for the collection of literature as well as for the newsletter narrowed to only LEISA related material. This face ended with the publication of a handbook on LEISA: “Farming for the future”, edited by the ILEIA team, published by a non-commercial publisher. In this book the secondary documentation was brought together. The reference list of this book gives a good impression of the grey literature collection of ILEIA at that moment. This publication was successful. Originally published in English, it was translated into Spanish, Portuguese, French, Arabic, Chinese, and Bahasa. It was distributed to a large audience.

One of the factors that contributed to the capacity of the magazine was the careful management of information entering the organization. The regular selection of material and the review of documents by ILEIA's documentalists and editors supported by external experts and specialists when necessary was a key element in this process. In addition ILEIA documentalists took an active part in European initiatives to establish LEISA documentation in academic and research collections in both Northern and Southern countries.

Between 1991 and 1999, the magazine began to systematically synthesize and analyze LEISA. Concepts were developed further and the network strengthened. The network and the magazine were increasingly used in the production of secondary published material.

### **ILEIADOC**

At about the same time the collection was subject to comprehensive cataloging, a procedure of developing key words and abstracts for the documents in the collection that is in line with standard practices and which continues to be refined and elaborated. A LEISA thesaurus was developed to facilitate retrieval of information. Metadata of the library collection were stored in a database known as ILEIADOC, this was initially made available on disc, and later could be accessed via the Wageningen Agricultural University on ILEIA's website. ILEIADOC was one way in which knowledge about the information available within ILEIA entered the public domain, and the increased awareness of the nature and scope of its LEISA collection also enhanced the status of the grey literature it contained. Indexes were regularly published and bibliographies were compiled which in themselves added to the retrievability of material being published and collected.

LEISA became a known alternative for expensive and unsustainable high input agriculture. More literature became available, also commercially published documents were received and included in the collection. The thematic concept for the magazine was developed further and with the themes for different issues the acquisition of thematic secondary documents started. Selection criteria for acquisition narrowed towards the themes that were under construction. Each theme highlights an aspect of LEISA but from a different angle. Some themes are repeated regularly like soil fertility for instance, others are dealt with only once or twice. Consistent search for this material resulted in a well-known and established collection that attracts a wide variety of material both grey and from the commercial press publishing development material.

The success of the book "Farming for the future" contributed to the development of a strong network. This network and the well-known collection attracted enough material to produce an interesting Magazine. The articles of the magazine, often written by field practitioners of LEISA methods, became important new documents on LEISA techniques themselves. With these articles, primary documentation entered the documentation collection of ILEIA.

### **From information to knowledge**

ILEIA occupies a special place in the process of creating knowledge from grey literature and other material on LEISA. During a highly specialized editorial process author contributions as well as primary and secondary literature and information confirmed and validated by network partners are forged together within the body of the magazine to make a coherent and comprehensive body of knowledge capable of elaborating the theme of each the edition.

After 1999 primary documentation became increasingly important. Documenting experiences and cases where LEISA was practiced has become the main objective of the magazine. The work of ILEIA has been concentrated on the dissemination of practical experiences with LEISA. ILEIA seeks to be a link that connects local level experiences to local level experiences elsewhere, providing a platform for sharing of information and learning from experiences of different countries, regions and continents. This formula proved to be successful, at the moment LEISA Magazine has 20,000 subscribers in 173 countries.

### **Extending the process**

The process of systematically synthesizing and analyzing LEISA has taken on a new dimension with the establishment of regional editions of the magazines. In Latin America, Indonesia, India and West Africa the production process is the same as that used in developing each edition of the global LEISA magazine. Regional editions make it possible to increase the information network available at grassroots level. It also requires care and coordination to secure reliable and relevant information. This is achieved by close cooperation and regular meetings and consultations between the six editorial teams. Until the establishment of the Spanish French and Indonesian editions of the LEISA magazine, all communications and publications had been in English.

Currently, ILEIA's focus remains the documentation of field experiences and therefore primary documents and grey literature continue to be an important part of the collection. The LEISA magazine and the LEISA Website are the main vehicles of information exchange, but publications with case studies and a CD ROM with all the magazine articles are in preparation. This together with the digital developments of recent years changed the function of the documentation collection (ILEIADOC) from one of maintaining a collection to an active sharing of the information it contained. The collection itself generated interest by other parties like Wageningen University and Research Centre (WUR). A part of the collection, the Magazine articles, is available via WUR databases. The collection as it is now contains 8000 books, reports, articles, CD ROMs and video recordings, and 150 journals related to LEISA.

### **Digital future**

In the near future the collection will develop into a digital collection. A new Website and new databases are under construction. This development will make it possible for ILEIA to make primary documents in full text available from the collection. The future of the documentation collection will be directed towards web-based primary documentation and secondary documents made available online as pdf files and xml fragments or links to pdf files and xml fragments.

### **Strength and weaknesses of the LEISA collection**

Within the field of development-orientated organizations, ILEIA has defined a clear target group of LEISA practitioners and extension workers at field level. This strengthened the content of the magazine and the content of the documentation collection. By providing strong and applicable information at field level, ILEIA established a reputation for reliability and a strong network of organizations and individual field workers. This in itself added value to the collection. The network generates the important input of articles for the magazine, and facilitates the primary documentation process. It contributes to the collection of documents and also forms one of the most important user groups of the collected material. The LEISA collection is a special collection with its own audience.

The disadvantage of being specialized is that completeness and maintenance cannot be the objectives for the collection. Selection criteria change continually and collection development depends for a large part on the received material, which means arbitrariness. The most recent material is the most interesting and older material needs to be reviewed for relevance regularly. Collection criteria are developed in the first place for internal use by the ILEIA team. For users outside the ILEIA team a lot of interesting often "grey" material about LEISA methods is available but the collection does not try to answer all questions. The collection lacks a number of expensive basic (not grey) manuals.

The facilitating environment provided by donor funding makes it possible to produce LEISA magazine with thousands of free subscriptions in developing countries. This strength implies also a weakness. The availability of donor funding determines the possibilities for acquisition of material. Continuity is not guaranteed and maintaining a special collection, though expensive is often not valued by donor organizations.

## Conclusions

Collecting, collating, presenting and disseminating information are key success areas for information centers. This is where many organisations either gain or lose competitive advantage. Information not managed professionally and constructively has little value. It remains inaccessible, and its existence unknown. In this sense ILEIA has profiled itself in such a way that it has maintained its capacity to continue the strategic collection of grey literature and to retain a status which has ensured and is recognized as having a special niche in development-orientated publishing and information exchange. ILEIA has been successful in collecting and disseminating information during the 20 years of its existence. The main reasons for this success must be sought in the methodology underlying the editorial approach and the continuous attention that has been given to networking and profiling ILEIA's activities.

## References

- Farming for the future: an introduction to Low-External-Input and Sustainable Agriculture by Reijntjes C, Haverkort B and Waters-Bayer A, 1992, London: MacMillan Education, ISBN 0 333 57011 1.
- LEISA Magazine: quarterly journal by ILEIA, Amersfoort, NL
- LEISA Revista de agroecologia by ETC Andes and ILEIA, Lima, Peru
- LEISA India: quarterly journal by AME and ILEIA, Bangalore, India
- AGRIDAPE Agriculture durable a faible apports externes: quarterly journal by IIED, Senegal
- SALAM Majalah pertanian berkelanjutan: LEISA Indonesia by VECO, Bali, Indonesia

---

## Author Information

### Wilma Roem

ILEIA, library  
P.O.Box 2067  
3800 CB Amersfoort  
The Netherlands

w.roem@ileia.nl  
m.minderhoud@ileia.nl  
www.ileia.nl



## The Information Market for Research and Higher Education How to integrate all relevant information in a network of repositories?

Hans E. Roosendaal  
University of Twente, UTwente

### Abstract

*The information market for research and Higher Education (HE) will in future be based on a federated network of repositories of information relating to research and education that conform to open standards, and an accommodating infrastructure that allows users the easiest and fastest possible access to information in all of these repositories.*

*The information covered by such a network will not only comprise of information material for research and HE, including grey information, but also of management information relating to this information.*

*The market is the research and HE community; its main focus is open systems.*

*This federated network will be global.*

This vision of a federated network of repositories of information for research and higher education will be analysed from a strategic point of view, with emphasis on the consequences for grey literature. In particular, we will discuss engines for change in this market and availability versus actual use of information is used as an important parameter. The value chain is used as analytical instrument to discuss a number of options and the roles and responsibilities different stakeholders will likely assume in these options.

The potential of a high level strategy for this market and its consequences for the different stakeholders in the value chain, and in particular the research and higher education institutions, is analysed.

These considerations lead into a strategic agenda for the research and HE institutions for all relevant information they are using and producing, in research and e-learning, in terms of content and management of this content, also for grey information.

### Justification

This paper is a record of an invited presentation given at the Fifth International Conference on Grey Literature held December 4-5, 2003 in Amsterdam. As such, the paper cites freely from a previous publication<sup>1</sup> of the author, albeit sometimes in a different context resulting in some different and new observations and conclusions. For general references the reader is referred to the references mentioned in <sup>2</sup>.

---

<sup>1</sup> Hans E. Roosendaal, "Driving Change in the Research and HE Information Market", to be published in Learned Publishing

<sup>2</sup> Hans E. Roosendaal, Peter A. Th. M. Geurts and Paul E. van der Vet, "Developments in scientific communication: Considerations on the value chain". Information Services and Use, 21 (2001) 13-32

## 1 Introduction

*'Authors want to publish more, readers want to read less.'*<sup>2, 3</sup>

This statement paraphrases the fact that wide exposure is paramount to the author and (pre)selection to the reader of research information, including grey information. Any force in the market like the use of Information and Communication Technology (ICT) by the actors involved (authors, readers, libraries, scientific publishers etc.) that allows better fulfilling this statement is an engine for change in the value chain, prompting changes in the roles of the stakeholders in scientific communication.

The above statement means that, for the author, visibility is crucial whilst, for the reader, retrievability is. In this context it is important to bear in mind that readers, when searching for information, will in most cases not be able to specify in detail what they are looking for. Combining these various factors can only lead to the conclusion that wide availability of information is the foremost requirement in this market. Arguing along the familiar business criteria of volume and margin we see that wide availability takes the role of high volume and restricted availability that of low volume. In the research and HE information market volume is thus the potential volume of readers, rather than the actual volume of reading. The fact that readers want to read less but everything that is relevant to them at the right time illustrates this point of view. This means that the elasticity in the market is determined by the degree of availability, and this is compatible with the requirements for an open system.

This discussion illustrates that the statement at the head of this introduction determines to a large extent the dynamics of the market, and is independent of the carrier of the information, be this paper or a digital carrier. In other words, the value chain of the research and HE information market is largely determined by it. In this value chain the author and the reader, jointly the user, are the generic stakeholders while other stakeholders are institutional stakeholders.

## 2 Vision

*The information market for research and Higher Education (HE) will in future be based on a federated network of repositories of information relating to research and education that conform to open standards, and an accommodating infrastructure that allows users the easiest and fastest possible access to information in all of these repositories.*

*The information covered by such a network will not only comprise of information material for research and HE, including grey information, but also of management information relating to this information.*

*The market is the research and HE community; its main focus is open systems.*

*This federated network will be global.*

This generally shared vision describes a real network of repositories of information relating to research and education, containing both research and education information in the widest sense and management information to support access to and disclosure of this information. The user, be they student, teacher or researcher, will be able to make use of this information from any location, at any time and in all possible ways. Realisation of the vision is not considered so much

---

<sup>3</sup> Hans E. Roosendaal, Peter A. Th. M. Geurts and Paul van der Vet, "Higher education needs may determine the future of scientific e-publishing". [www.nature.com/nature.debatess/e-access/Articles/roosendaal.html](http://www.nature.com/nature.debatess/e-access/Articles/roosendaal.html)

a technical issue, but an organisational issue: how do we get the transformation done? In particular the fact that proprietary is not anymore sought means a real transformation in the mind of the stakeholders as the so familiar paper environment is very proprietary indeed. The system should be open: this is not only a requirement to the software, but also a requirement to the concept of the entire system. This allows convergence of white and grey literature. These are the real changes in the value chain.

### 3 Engines for change

Changes in the value chain are triggered by engines of change. For this market these engines for change are the potential that ICT offers to empower the author and reader and the recent developments in research and HE, also to a large extent but not exclusively enabled by the potential offered by ICT. ICT provides a huge potential to empower the author and the reader and allows a change from a use-oriented system towards a more availability-oriented system at the same time allowing a new balance between centralised systems and distributed or federated systems. ICT raises for the stakeholders the strategic choice between empowerment of the user, or alternatively applying a hostage strategy<sup>4</sup> directed at the user in particular.

With respect to some broader developments in research it may suffice to mention that research has generally become more subject to market conditions, even when carried out in the environment of a research institution. Market conditions mean that intellectual capital and scarcity of resources, both financial and human, play a more and more important role. As a result, research information is being intensively used for planning and evaluating of entire research programmes emphasising the formal publication side system rather than the communication side. This means to say that the balance between real communication between researchers as opposed to formal publication of research information is even more changing to formal publication.

In education, the introduction of the bachelor/master structure at the European universities will spur the development of web-based and blended learning when students are becoming more mobile and will hop from one university to another. This mobility is expected to show up in particular for master students and will lead to the introduction of international masters. Wider applications of distance learning and life-long learning will spur these developments.

For our discussion it is interesting to note that the information requirements - in terms of publishing and archiving - for research and for educational materials are very similar indeed. For educational information the volume required for each HE institution is at least an order of magnitude larger than the research information it requires. This makes it attractive from an institutional point of view to have research information financially piggy-backing on educational information. HE institutions have to develop their information infrastructure for the production and registration, i.e. publishing and archiving of educational material anyway and can use that infrastructure for the production and registration, i.e. publishing and archiving of research information as well. In both cases this includes the production and registration, i.e. publishing and archiving of grey information.

---

<sup>4</sup> Edward Freeman & Jeanne Liedtka, "Stakeholder Capitalism and the Value Chain". *European Management Journal*, 15 (1997), # 3, 286-296

#### 4 Functions in scientific information

The main functions in scientific, research and educational, information is defined in<sup>2</sup> as in Figure 1 below.

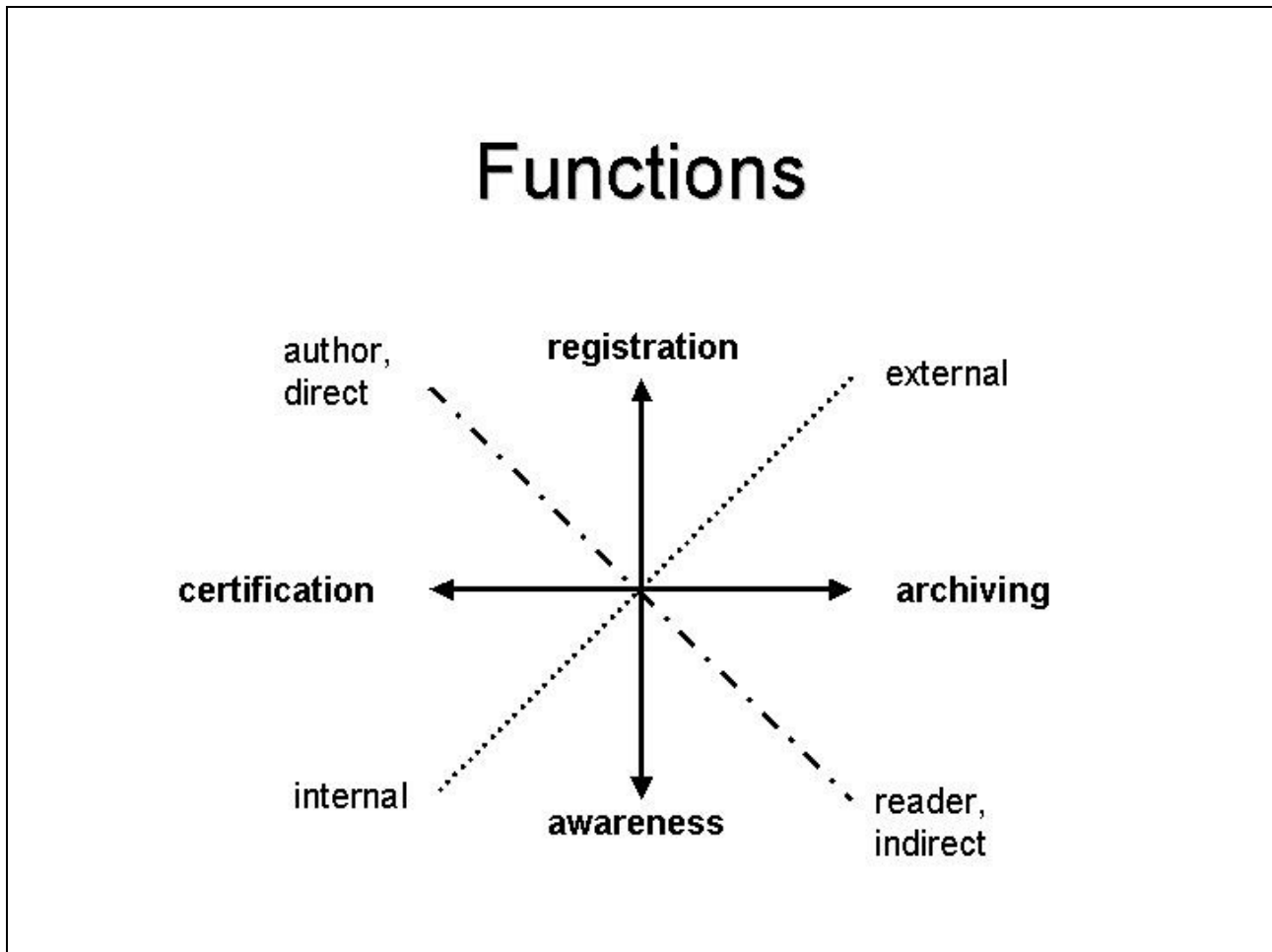


Figure 1: strategic functions of scientific information

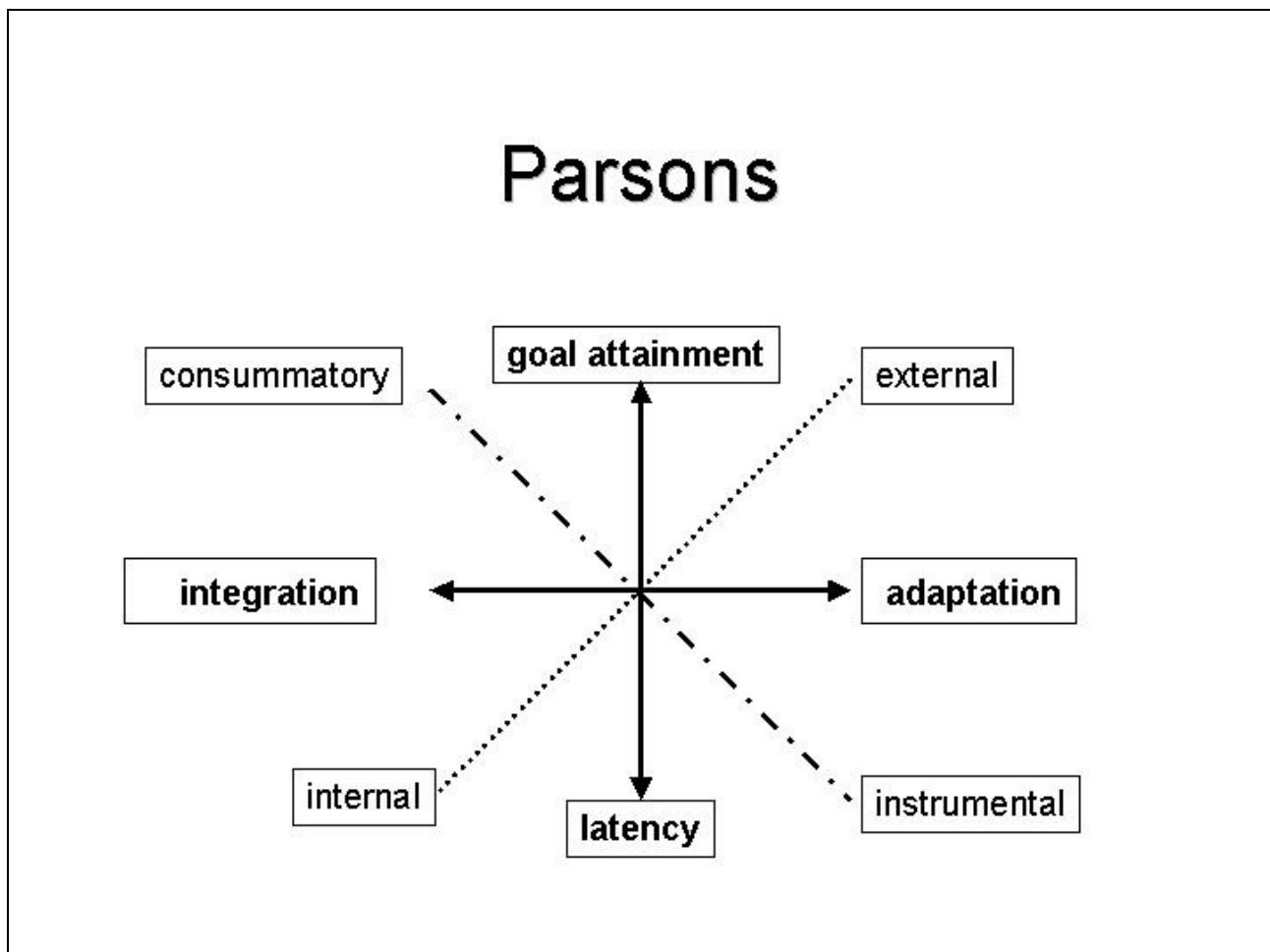
These functions are defined as strategic functions from a science point of view. The external functions registration and archiving are seen to be outsourced out of science to the publisher and the library respectively. This four function scheme serves as a starting point for strategic considerations, as e.g. it has been used for e-publishing in the OpenArchiveinitiative strategies<sup>5</sup>.

This four-function scheme has a longstanding tradition, see e.g.<sup>6</sup>. Characteristic is the 2x2 classification of external/internal and concrete/abstract or direct/indirect dimensions. This 2x2 classification has also inspired Parsons in his seminal theory<sup>7</sup> on a social action system as shown in Figure 2.

<sup>5</sup> see e.g.: Thomas Krichler and Simeon M. Warner, "Open Archives and Free Online Scholarship", Proceedings of JCDL2002

<sup>6</sup> Harald Atmanspacher, "Raum, Zeit und psychische Funktionen", in H. Atmanspacher, H. Primas, E. Wertenschlag-Birkhäuser (Editors), "Der Pauli-Jung Dialog", Springer-Verlag, Berlin, Heidelberg, New York, 1995

<sup>7</sup> Talcott Parsons, "The Structure of Social Action", The Glencoe Press, Glencoe, 1949 (1937)



*Figure 2: 4-function scheme of Parsons<sup>7</sup>*

Again we see that the elements of social action or transformation are classified along the external/internal and in this case consummatory/instrumental dimensions. The four function scheme is thus seen to be firmly embedded in our scientific tradition, and in different scientific areas.

The four functions in scientific information need always be performed independently of the technological environment, albeit that the balance between the functions may well change under changing technological conditions. We will use the four function scheme for scientific information as analytical tool in our analysis of changes in the value chain arising from the engines of change as discussed above. They provide amongst others a powerful check on the comprehensiveness of these scenarios and the consequences for the stakeholders.

## 5 Some observations and considerations

One may argue that the vision as presented in Chapter 2 is not really new as the invention of the journal in the seventeenth century is already an early expression of a generally available repository of information making use of the then existing technology. The present, strategic consequence of institutional repositories is that they will lead to combining the two strategic functions of registration and archiving into one hand, i.e. the institution of the author of the work. This is a new and important strategic development having its consequences for the value chain. And this again means that the distinction between white and grey information may shift to a different level, as we have noted before.

Presently, research and HE institutions and other knowledge-intensive organisations and companies world wide are developing novel but often disparate approaches to creating institutional repositories. HE institutions will have to create institutional repositories if they want to embark on e-learning, meaning that this is an autonomous development for these institutions. The network will connect the existing and future repositories and libraries. This requires a coherent user access or a shared information environment.

To achieve wide acceptance of the network it must contain a sufficiently large critical mass of information material. Critical mass is also needed to be able to support a variety of value chains for the information market. These different value chains should allow different organisational models, legal models and business models as the individual stakeholders see fit for the exchange of specific information products.

The creation of a cohesive and coherent network guarantees the best return on investment, in whatever form, for all stakeholders on their own terms, be they public (such as e.g. universities) or private (such as publishers) organisations, and whatever currency of exchange<sup>8</sup>, they prefer to use: money, power, influence or value, or combinations of these currencies.

It is therefore in the interest of each individual stakeholder to strive for maximum flexibility in the market place. This can best be achieved by developing a strategy that allows maximum compliance with the vision in the market place. Stakeholders should share a basic conception of such a high level strategy as strategic instrumentation for providing a starting point for developing their individual strategies. This strategy will be briefly discussed in section 7. In the next section a number of value chain options and their consequences will be discussed.

## 6 Value chain options

As stated above, ICT in particular allows a variety of value chains. The value chain is defined being linear in terms of steps of added value and is not a process chain. The corresponding process chain is in essence a rather complex network of process steps.

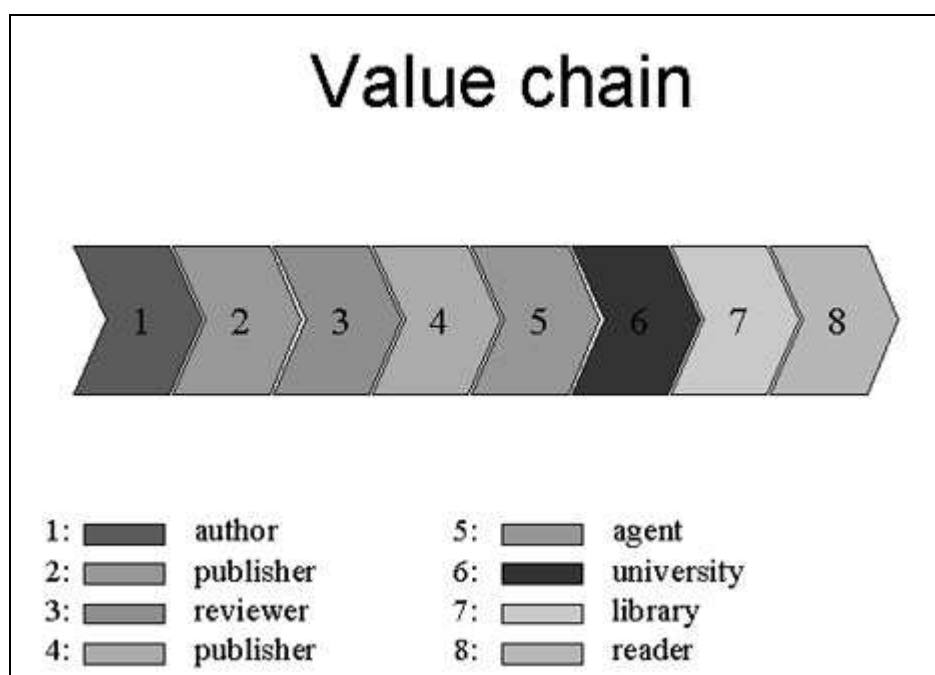


Figure 3: traditional value chain

<sup>8</sup> Talcott Parsons, "Evolutionary Universals in Society", in "Sociological Theory and Modern Society", The Free Press, New York and London, 1967 (1964)

Figure 3 shows the traditional value chain, as we know it from the paper-based environment. In this figure we show the value chain with the stakeholders responsible for the added value per link. Thus the author creates the work, sends it to the editor, the publisher will produce the work and send it to the university. Administrative assistance is mostly given by an agent. Finally the paper arrives at the reader.

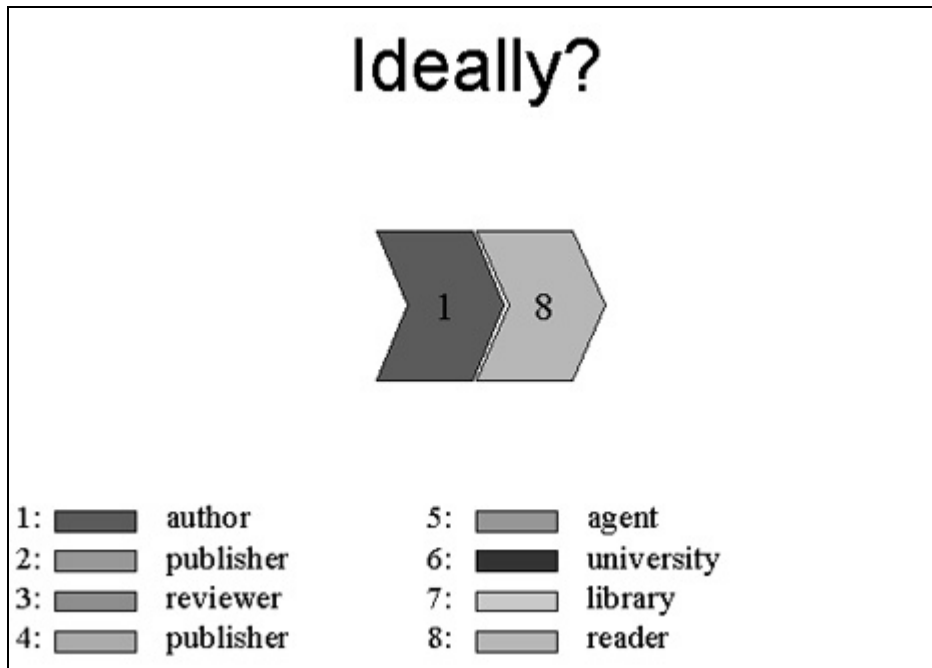


Figure 4: value chain with full empowerment for author and reader

In figure 4 we show a shortened value chain of author and reader only, i.e. full empowerment for the author and the reader. This means no quality filter or branding. This value chain can well work for information that the reader is very familiar with, but takes an extraordinary effort on the part of the reader with information less familiar thereby violating the statement: *'Authors want to publish more, readers want to read less.'* This value chain is totally availability-based meaning that the author or the institution does not only have to bear the financial risk but as there is no refereeing there is also for the author the full risk as scientific entrepreneur.

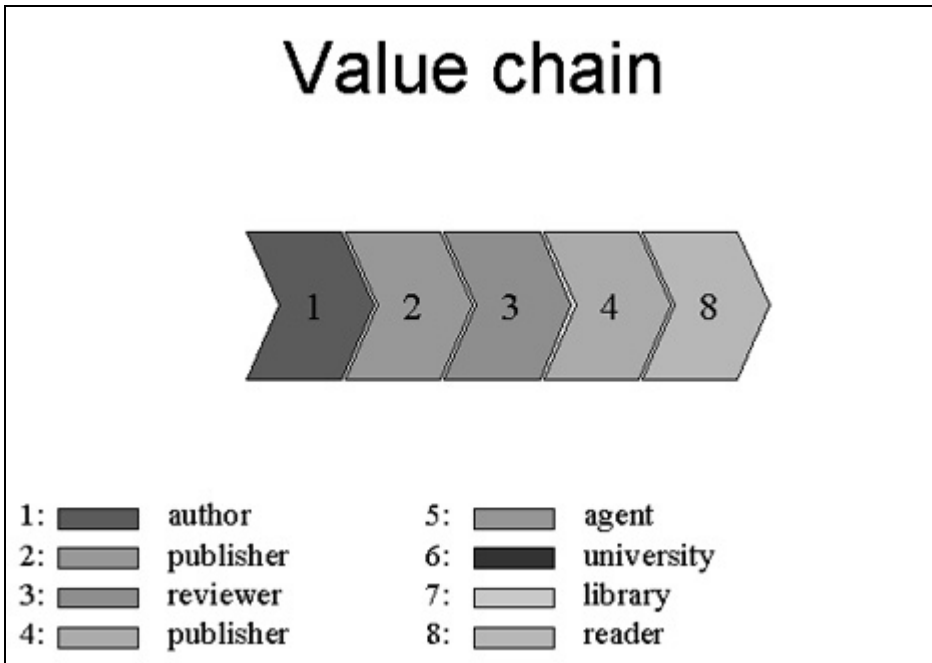


Figure 5: value chain without universities

Another possible value chain is the one in figure 5 where publishers are delivering information directly to the reader. Weak point in this value chain is the responsibility for the archive that in this case should rest with the publisher, not a very realistic proposition. This value chain is totally reading use based and costs will have to be picked up by the reader.

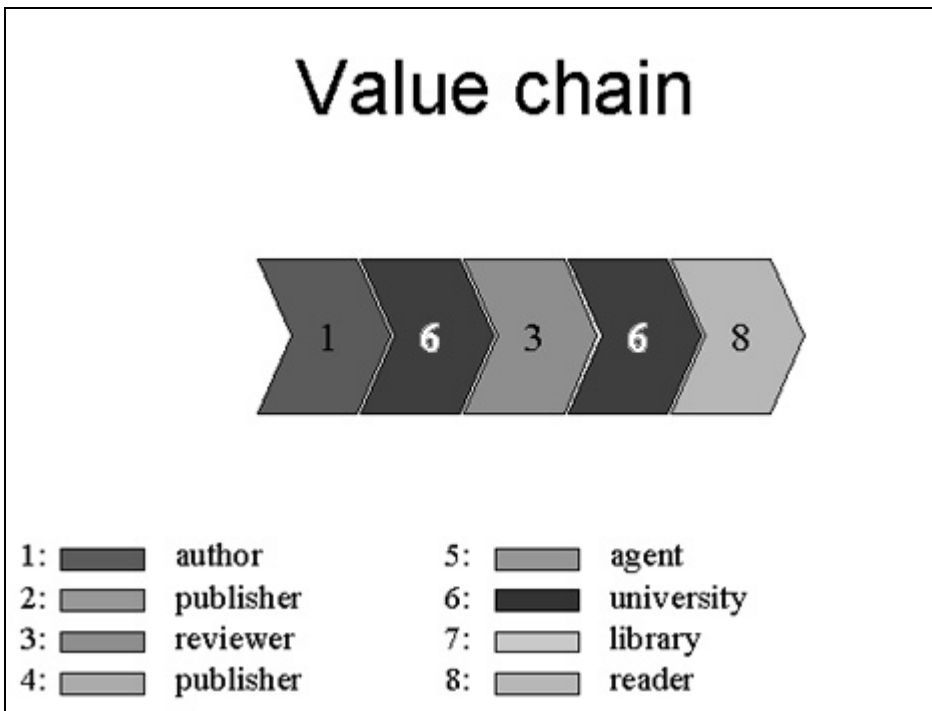


Figure 6: value chain without publishers



Alternatively, we could swap the publishers for the HE institutions taking over the publishing function. In the case of research information the weak point then is the certification of the material. This cannot be managed by the home institution of the author. A way out could well be the creation of alliances of institutions, leading finally to the establishment of new publishers. However, for learning material this value chain is highly feasible as in this case the 'buying' institution can exercise the certification power.

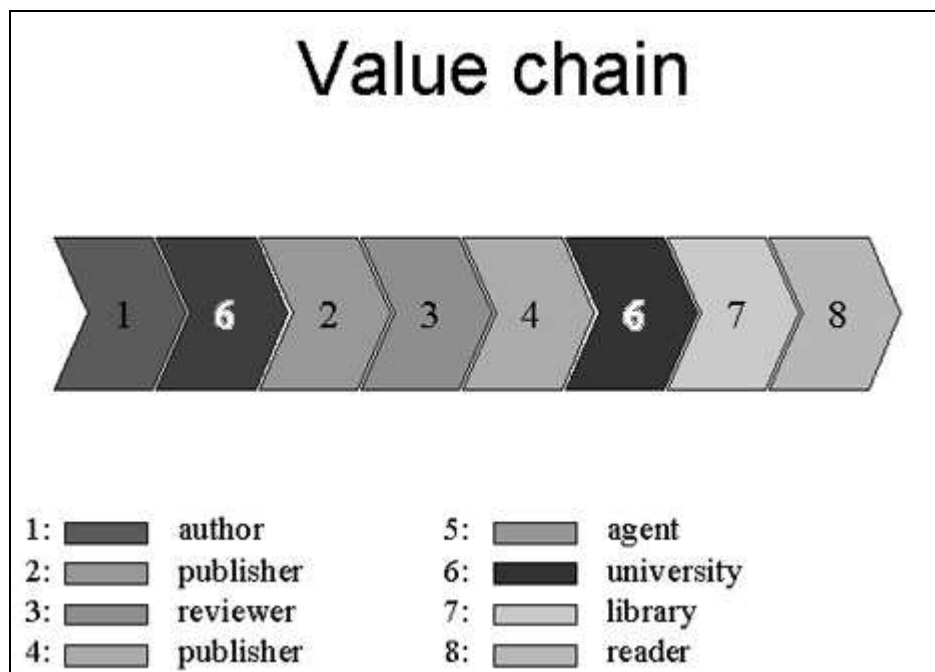


Figure 7: value chain with new roles for institutions and publishers

In the last figure we see a value chain that looks rather similar to the traditional value chain, but with totally new roles for the stakeholders. The institutions are now responsible that the work (author) can be sustainably archived ('perpetual' archiving) and is properly disseminated to the reader. The institutions are in this chain responsible for the registration and the archiving functions<sup>2</sup>. The publisher is responsible for the distribution and branding and in providing logistical assistance for the editor in the certification process.

This chain has a number of consequences:

1. the fixed first copy costs have not to be born twice, i.e. by the institution and the publisher, but will be born by the institution only;
2. the author can transfer copyright, i. e. the overall exploitation rights, to the institution where the work was performed. The institution can then transfer specifically designated rights to the publisher.

In this last value chain the costs for production and dissemination will be born by the institution. The different options of the value chain allow different options for scenarios in the research and HE information market. The different options represent also differences in the balance between availability on the one-hand side and reading use on the other hand. This is relevant for the different business models emerging from these options.

## 7 The high level strategy

Creating a network of repositories of information relating to research and education requires a basic conception of a high level strategy shared between the different stakeholders having different business philosophies. Such a strategy can only be successful if it fulfils in the best possible way the major interests of the stakeholders. This requirement means that such a strategy can only have one focus: the user as the primary beneficiary of the network. This is the only possible strategy leading to value creation, the alternative being value capture by one of the stakeholders and taking the other stakeholders, in particular the users, as hostages<sup>4</sup>. The user is the learners, teachers, researchers and students in knowledge institutions and organisations, in their capacities as author and/or reader. This means that a comprehensive approach to user behaviour and to the consequences of such behaviour for the value chain of information is indispensable.

The institutional stakeholders in the research and HE information market and beyond, will as enablers be the secondary beneficiaries. As stated before, the foremost goal for every stakeholder is to develop an individually tailored strategy to comply with the high level strategy in this way positioning this stakeholder at the forefront of developments in on-line information management. Only then the stakeholder will be able to make an invaluable contribution to a network for world wide information provision in research and education. A key aim of this strategy is making universities and other knowledge institutions, scientific publishers, non-commercial or commercial, professional by helping to make use of this network and ensuring that the architecture will best serve all stakeholders' needs.

The network should be able to support the user in the strategic tasks as embedded in the value chain in Figure 8.

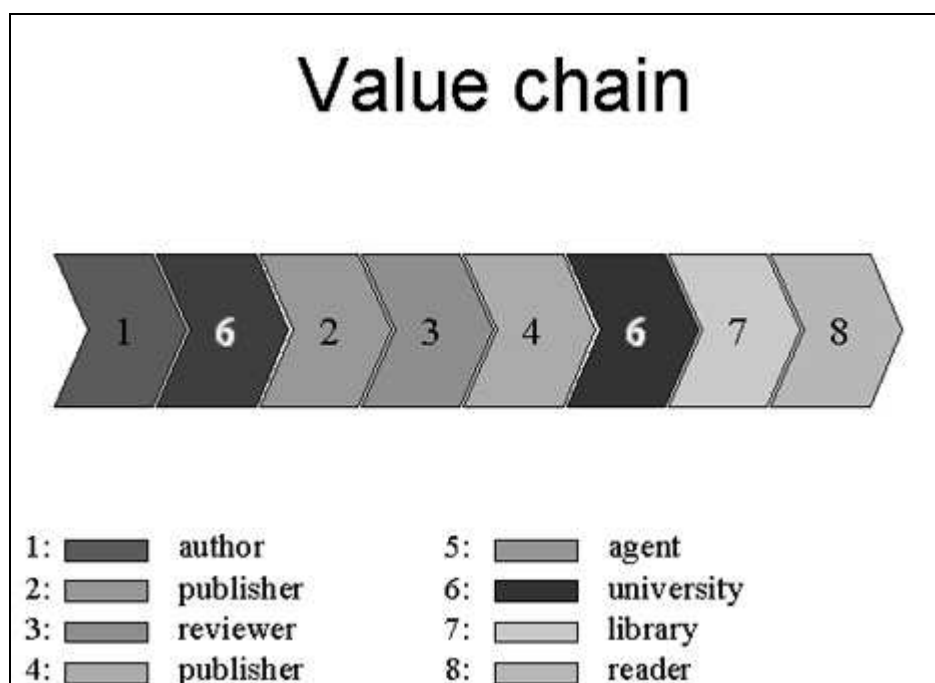


Figure 8: strategic tasks of the network as represented in the value chain

As a consequence of such a high level strategy the corresponding technology strategy should focus on developing an architecture for federating existing and future repositories and libraries for the familiar strategic reasons for making use of an architecture<sup>4</sup>:

- to reduce complexity;
- to allow a proper balance between central and decentral aspects of the development;
- to be able to manage change properly;
- to facilitate experimentation and competition;
- to ensure that many different systems can develop together gracefully.

A main goal for this architecture is the development of a shared architecture for e-documents, e-learning and e-science and this requires integration and resources syndication. A foremost strategic goal is that the authentic copy of a work of whatever type, should remain located at the home repository, being the repository of the affiliation of the creator of this work. This would constitute an important step towards empowerment of the user.

## 8 Consequences for stakeholders

A network of repositories leads to a transformation of the value chain and different roles of the stakeholders in this chain. The most important change is the anticipated splitting of the publishing function: the separation of the registration function from assistance in the certification function. The registration function will then be combined with the archiving function and these two functions will probably become the responsibility of the research and HE institutions. This leads to a substantial change in the position of these institutions in the value chain. This will support these institutions in further developing into institutions exchanging, under business conditions to be agreed, educational content between them, and to become as institutions creators of specialised educational content in their own fields of expertise.

Under these developments, the function of the publishers will become much more a facilitating and aggregating function where distribution and dissemination will be replaced by providing access, primarily to certified content available in the repositories. There are ample opportunities for forward and backward integration and this may well be attractive to existing publishers in view of the large volume, or rather availability, of this content.

Transforming towards a federated network of repositories means transforming towards a federated ownership. This development is starting just only now and much more thought is needed to further it. However, this issue should be solved as it otherwise becomes a real impediment to this development of the market. Shared vision and value chain considerations are required to be able to develop business and distribution models that will be sustainable for some time to come.

The issue of federated ownership is tightly related to the issue of trust relations between the stakeholders. Therefore, these trust relations are also a major issue for research<sup>9</sup> and education as well.

## 9 Concluding remarks

For further developing the research and HE information market in a digital environment it is highly desirable to develop and test new and different value chain models, and consequently business and distribution models. We have seen that institutional repositories are anticipated to

---

<sup>9</sup> Harry Hummels and Hans E. Roosendaal, *Journal of Business Ethics*, 34 (2001) 87 - 100

lead to recombining the registration and archiving function at the research and HE institution. This development makes the research and HE institutions the natural candidates to initiate the development of new value chains and new models. Although this development is a formidable technical task, the issue is not so much a technical, but rather an organisational one. This development bears important consequences for the strategic development and use of grey information. Rather than seeing grey literature as type of product or a set of types of products it may well be tempting to consider grey literature as a specific type of value chain or a set of specific types of value chains in the entire family of value chains possible in information related to science. Indeed, in grey literature the registration and archiving function have always been combined at the author's institution, being this an individual author or the institution itself. Grey information that is published on the institutional repository will then enjoy wide availability as opposed to limited distribution as used to be the case and this will make grey information right away the most abundantly available scientific information.

---

### **Author Information**

---

#### **Hans E. Roosendaal**

School of Business, Public Administration and Technology  
Department of Computer Science  
University of Twente, UTwente  
P.O. Box 217  
7500 AE Enschede  
The Netherlands

H.E.Roosendaal@utwente.nl  
www.utwente.nl

## **The moving border of tacit and explicit knowledge in e-Learning: Use and production of information and knowledge in technical university education, case study**

Anna-Kaarina Kairamo

Helsinki University of Technology, Teaching and Learning Development

### **ABSTRACT**

In this paper, the electronic information and knowledge created within a context of a course is approached by analysing two case study courses. The Finnish context of the university education is first outlined briefly. Approaches to learning resources are reusability and course components. The two case studies are a series of courses of energy engineering and environmental protection, and a programming course. The courses were developed in different department cultural context and circumstances, which are described. When the focus of course material creation is on methods, the reusable material might be related to resources meant to help the student.

### **INTRODUCTION**

One of the often mentioned benefits of e-Learning is the externalisation of the content and process of the courses, which can support reuse of material across courses. Yet most of the course management systems financially available for higher education organisations do not support ideas of knowledge transfer or learning organisation and content remains blocked to one course.

Libraries, information services and commercial organisations provide access to electronic materials. However, these possibilities do not seem to be widely used in higher education. There may be several reasons for this: complicated procedures of intellectual property rights, attitudinal and cultural barriers and even teachers' inadequate skills in searching the electronic information and knowledge. One fundamental reason might be that strict externalisation is not regarded as fruitful in knowledge transfer, or that the externalisation effort considered huge compared to the short lifetime of the information and knowledge. Also the available technology in the organisation may be hard to use or may not support the ideas of the teacher.

The OECD report on Knowledge Management in the Learning Society (2000, 12) claims that we have not yet reached a stage where we can systematically apply knowledge to the production of knowledge. Although higher education knowledge is highly codified, there is much tacit knowledge in teachers' know-how (Ibid 31).

Do we really know what is happening in grass-roots level of ICT enhanced higher education? What kind of information and knowledge is available for the whole community of a course, which includes instructors and students? What kind of information and knowledge is produced by the instructors and is there any knowledge used as a resource, created by the student? What kind of external resources are used? What is the potential reusability of the information and knowledge produced for the course? These are the main questions of this paper.

In the following, the national context of the university education is first outlined briefly, before going to the approaches to learning resources. The method of case study is then introduced and the cases described briefly. Finally, the first results of the study are presented and discussed.

### **CONTEXT: NETWORK-BASED EDUCATION AS A GOAL FOR UNIVERSITY SECTOR IN FINLAND**

The Finnish 'Development plan for education and university research for the period 1999-2004' states the following about university education:

In university education, the aim will be large-scale pedagogical renewal towards student-centred teaching methods. The development of teaching and learning will especially capitalise on network-based and open learning and distance learning.

As a result of the 'Information Strategy for Education and Research 2000-2004', the Finnish universities established in January 2001 a consortium of the Finnish Virtual University. Although FVU's role is not to provide academic education itself, the vision of FVU reflects the national network-based education policy for universities, agreed by the consortium representatives mainly consisting of university leaders:

By the year 2005, high-quality, ethically and economically sustainable modes of operation in network-based education and research will be in widespread use in Finnish universities.

The basic mission of Finnish universities is to carry out research and provide education based on it. The basic principles in university education are the freedom of research and university autonomy. For example, the strategies for the case organisation Helsinki University of Technology (HUT) do not state explicitly how widespread the network-based education at HUT should be. Thus it leaves, in tradition of autonomy, the concrete goal setting of the extent and form of use of the network to the departments, laboratories and teachers. The underlying interest of this study is get further information on what kind of models of e-learning solutions and knowledge creation in the context of higher education emerge in this kind of organisation.

### **HUT profile in change process**

In her research, Fisser (2001) constructs four possible profiles of universities that relate to the change process of using new forms of ICT in education. The profiles differ, according to the reason of the change (social vs. economic) and the experienced environmental context (uncertainty vs. interconnected):

Profile 1: Supporting learners in a changing world

Profile 2: Developing expertise in the institution

Profile 3: Surviving in a competitive world

Profile 4: Evolving to a cost-effective approach in education

HUT was one of the four case study universities of the research. Among the observations made of HUT in the research were the following: The primary motive to use ICT in education was social. The new forms of ICT have not brought a major change in the way education is offered and supported. ICT has been implemented in education on a voluntary basis. The suitable profiles of HUT were 1 'supporting learners in changing world' and 2 'developing expertise in the university' with the majority of the respondents in profile 1 (Fisser 2001, 148).

## **APPROACHES TO LEARNING RESOURCES**

### **Digital learning material, one special form of grey literature**

The Fourth International Conference on Grey Literature defined grey literature as information produced on all levels of organisations in electronic and print formats not controlled by commercial publishing. The rapid increase of use of ICT in education in all forms from distance education to ICT supported traditional teaching has increased the amount of networked information and knowledge in university sector. Most of this information is not commercial, and thus can be regarded as grey literature.

In this study I will focus on digital material produced within higher education courses.

## Reusability of learning resources

Reusability of learning resources is widely discussed area in education field (e.g. Littlejohn 2003, 2). The concept of learning object (LO) is in the core of this discussion. Koper (2001) defines learning object as any entity, digital or non-digital, that can be used or referenced during technology-supported learning. This is also the definition of the IEEE Learning Object Metadata standard. There are a number of attempts to narrow the scope of the definition. As Conole (2002, 10-11) points out, there are several unresolved issues in the learning object approach, such as degree of granularity. There seems to be mutual understanding that a reusable object is much smaller than a course. In general, the smaller the unit, the greater the possibility that it will be reused. On the other hand, larger object may include e.g. activities, and constructing the course from greater elements may save teacher's time. (Littlejohn, 2003, 4).

The most common, but often an implicit idea of a learning object is that it includes content and/or method (Koper 2001, 4-5). The empiric experience from engineering education field supports this analysis.

Bates et al. (2003) point out, that despite the rapid expansion of digital materials, it is often difficult to find materials of the desired quality or level to fit the planned teaching. Also the size and structure of the available material might not be appropriate for the purpose.

## Course components

There are many approaches to structure the course activities. Collis and Moonen (2001, 20-21) analyse a course in terms of components, each related to pedagogical activities involved. In the following their structure is modified for the purposes of this study, so that the digital information and knowledge of the analysed courses can be categorised:

### Course organisation

- course information
- administration: record keeping, student marks
- general planning for the course

### *Lectures and other forms of instructor-led sessions*

- lecture material, highlights etc.

### *Self study, assignments*

- readings
- activities and assignments
- practical exercises

### *Major assignment (project work, essay, product, case study etc.)*

### Testing and examination

### Guidance, mentoring

- guides, tips, etc.
- communication

The categories that might be potential for reusable digital resource are italicised in the list.

## E-LEARNING IN HELSINKI UNIVERSITY OF TECHNOLOGY - TWO CASES

### Gathering data for cases

The data for the two cases now in focus was gathered by interviews and process simulations sessions within the Helmi research project, which studies the innovations of the e-learning processes (Smeds 2003). The interviews and process simulation sessions made it possible to understand deeply the process of designing and implementing the courses and learning material within the context of different cultures and traditions of two HUT departments. In Helmi project, the process to be simulated is first analyzed, modeled and visualized on the basis of the interviews. The simulation session is a facilitated semi-structured group discussion. The common

memory of the group participating is a process map, projected on the wall of the simulation room. (Smeds et al. 2001, 173-174.) The aim of the session is to create a shared understanding of the process among the participants, who are the key persons of the process.

After the simulation sessions, the next step of the research will be a second round of analysis of the data collected from the case studies and the simulations. On the basis of further literature review and this analysis, 10 further cases will be chosen for further research.

### **Case 1: Energy engineering and environmental protection courses**

An e-learning project within the laboratory specialised in energy engineering and environmental protection began in 2001. The purpose of this e-learning project is to explore how ICT could be implemented for the education offered by the laboratory. The research case was a series of three courses offered mainly for students who have studied at least three years. The number of participants of courses varied between 12 and 20. The learning material produced during the project was awarded as the material of the year 2002 in the university.

In parallel with applying ICT in the education, new pedagogical ideas were applied, such as problem based learning sessions and giving credit points from assignments, instead of only valuing the performance of the student in examination. The project team wanted to test their new ideas with the technology. Already in the beginning of the project, the exploitation of the hardware and software provided by the IT centre of the university was felt too slow and restricted process. It was also not possible to use any interactive elements. The solution was to make the laboratory as independent as possible of the services of the main university. A temporary solution for this was to choose for course process use a course management system (CMS). The basic reason was to provide a common easy-to-use tool for a group of instructors and a shared workspace for students. The course management tool was abandoned after two years for several reasons: too high expenses, little added value compared to open web page system for a small course, the creation of the laboratory's publishing process on its own server, and the students preferences to use email for communication instead of the tools of the system.

### **Case 2: Computer programming course**

The case laboratory is specialised in information processing science, and it is responsible for the basic education of computer programming for all students of HUT. The research case is an introductory computer programming course meant for all students at HUT. The participants are mainly second or third year students. Every year around 300 students complete the course.

The current mode of the course is a result of an evolutionary process during the last ten years. The pedagogical idea behind of the course structure is that the only way to learn programming is practicing. Using ICT tools is thus natural; it is a content of the course. The challenge of the case course is the large number of students. The solution to this is to make the process of studying and the process of teaching and guidance as explicit and strict as possible. The course web site consists of tens of pages of information and hundreds of pages of learning material on the web. All this is accessible for all, only the software system used for accessing and submitting exercises is password secured. The staff includes the responsible teacher and assistant and around 20 part time assisting students. The community of these instructors works virtually: they meet twice during the half-year course, before the course start and after it has finished.

### **First results**

The digital information available for the whole community of the case courses (instructors and students) is categorised in table 1 using a modified model of Collis and Moonen (2001). The information available only for instructors, such as exercise products of the students, is not of interest, since the underlying assumption is that this private information has no potential for



reusability for training. The producer of the information is marked in bold in the table if it is someone else than the instructors.

Table 1: Digital material components of the case course. The materials of the case courses potentially reusable are italicized.

Digital information component	Case 1: energy engineering and environmental protection courses	Case 2: programming course
Course organisation <ul style="list-style-type: none"> <li>- course brochure</li> <li>- administration: record keeping, student marks</li> <li>- general planning and running of the course</li> </ul>	Basic information on time, place, requirements  Results of the assignments Final marks	Detailed information on the course arrangements, goals, teaching methods, course staff, creditation of components of the course, and course communication Results of the assignments Final marks Detailed instructions for assistants
Guidance, mentoring <ul style="list-style-type: none"> <li>- tips, etc.</li> <li>- communication</li> </ul>	Announcements during the course	FAQ (Frequently asked questions) <b>Instructors:</b> answering questions in the news discussion group <b>Students:</b> answering each others' questions in the news discussion group
Lectures and other forms of instructor-led sessions: lecture material, highlights etc.	<i>Lecture slides</i>	<i>Lecture material</i>
Self study, assignments <ul style="list-style-type: none"> <li>- readings, obligatory</li> <li>- readings, voluntary</li> <li>- assignments, practical exercises, other activities</li> </ul>	<b>Instructors:</b> Information on paper based reading, and where it can be reached <i>E-book available on www, cd-rom and printed</i> <i>Translation on part of a paper based book</i> <b>Students:</b> <i>Group assignment reports</i> <b>External resources:</b> E-book on cd-rom  <i>Equation collection</i>  <i>Exercises and model answers</i>	Information on paper based reading  <b>Instructors:</b> <i>Handouts of the course</i> <i>Information on additional resources, like books, web links</i> <b>External resources:</b> Web resources linked from the course site Guides and tips for programming
Major assignment: project work, essay, product, case study etc	<i>Group assignment instructions and background material</i>	Project assignment instructions
Testing and examination	<i>Pre-test</i>	<i>Previous examination questions</i> <i>Tips for examination</i>

The categories shown in table 1 can be clustered into those including content information, and those including method type of information, using the idea of learning object described in chapter 'Reusability of learning resources'. Method information includes categories of course organisation and guidance. The rest of the categories may be considered as content type of information. This classification is only rough and mainly clustering in its type: each category may include both content and method information. In case 1, the content type of information was

published as pdf-files and in case 2 programming course they were html-pages included in the web entity of the course and following the layout of it.

The information on the case courses differ in their completeness. In case 2 the student can find every detail of necessary information on web. In energy engineering and environmental protection courses case the web includes some basic information on the course arrangements. The focus of case 1 is clearly on publishing e-books and some support materials on web, thus on producing content type of information. In case 2 the focus is on computer based exercises and logistics and use of special tailored tool. In this case electronic information is highly method-type in its nature and thus may be more bounded to one course context.

Interesting question is the role of information that the students create during the course. In case 1 the written assignments acted as readings for the course. In case 2 the students acted openly and network-based as peer-supporters in newsgroup open for all. External digital resource of the case 1 is e-book developed by the partner university and in case 2 information behind links.

In case 1, the key enablers of the development process were the previous experiences of the instructors on implementing ICT in training and high technical skills of the key person of the project. A positive and encouraging culture for testing new ideas was also critical success factor. The active partnership with a foreign university within the field of education encouraged staff's endeavours. The key obstacle during the process turned out to be the strict rules of the IT centre of the university. The structures of published information are nowadays based mainly on the ideas of the laboratory, which might vary from the structures of the neighbour laboratories, not to mention other departments. The production system is vulnerable to a certain extent in such a small laboratory, since it is highly dependent on the skills of one person.

Within the department of case 2, ICT in education is not a tool, it is part of culture. As a result for solving the problem of a enormous amount of participants, the course is a very well organised and documented with all possible help for the student in written form on web. The course process resembles 'lubricated' machine, where the role of each actor is well-defined and strictly rule-governed: the responsible teacher, the assistants and the students. The structure of electronic information has developed in an evolutionary process. The web site of the course is partly cumulative and partly consists of detailed information that has to be updated for each round of the course.

## **FURTHER RESEARCH**

In this paper, the electronic information and knowledge created within a context of a course was approached by analysing two case study courses. In the context of the university, where the application of ICT in education is not centrally directed, different structures and content types emerge and the accuracy and extent of information and knowledge vary. There is also room for innovative solutions. In these case courses, potentially reusable digital resources may be content type of materials like e-books, information on additional resources including links, pieces of readings, and possibly student assignment reports. When the focus of course material creation is on methods, the reusable material might be related to resources meant to help the student.

Lots of questions remain unanswered on this phase of the study, and thus are worth for further examination: To what extent do the instructors search for possible external resources and why do they include or leave out materials in their course? How are the already produced resources of the courses reused in other contexts? How do the granularity and independence of the information object enable reusability? What are the criteria of the instructors for quality of information? To what extent is electronic scientific information used in courses and how?

## **ACKNOWLEDGEMENTS**

This paper is written in the Helmi project (Holistic Development of e-Learning and Business Models) at SimLab, Helsinki University of Technology (<http://www.simlab.hut.fi/>). guided by

Riitta Smeds, Professor of Business Process Networks and Director of SimLab. The author is grateful for the research collaboration of the multidisciplinary Helmi project team that has made this paper possible. The Helmi project financed by the Academy of Finland and its Life as Learning Research Programme, Tekes - National Technology Agency of Finland, the City of Espoo, Elisa, eWSOY, Laurea Polytechnic and Helsinki University of Technology, which is gratefully acknowledged.

## REFERENCES

- Bates, A.W., Poole, G. (2003). *Effective Teaching with Technology in Higher Education: Foundations for Success*. London: Jossey Bass.
- Collis, B., Moonen, J. (2001). *Flexible learning in a digital world: Experiences and expectations*. London: Kogan Page.
- Conole, G. (2002). Systematising learning and research information. *Journal of Interactive Media in Education*, 2002 (7): 1-28.
- Fisser, Petra (2001). *Using Information and Communication Technology. A Process of Change in Higher Education*. Enchede: Twente University Press.
- Knowledge Management in the Learning Society (2000). OECD report. Organisation for Economic Co-operation and Development.
- Koper, Rob (2001). *Modeling units of study from a pedagogical perspective. The pedagogical meta-model behind EML. First Draft, version 2*. Open University of the Netherlands.
- Littlejohn, A. (2003). Issues in reusing online resources. In Littlejohn, A. (ed), *Reusing online resources: a sustainable approach to e-learning* (pp. 1-9). London: Kogan Page.
- Smeds, R., Haho, P. & Forssén, M. (2001). *Implementing Knowledge into Action in Organizations: Simulation Games for Successful Process Innovation*. In Pantzar, E., Savolainen, R. & Tynjälä, P. (eds.): *In Search for a Human-centred Information Society. Reports of the Information Research Programme of the Academy of Finland 5*. Tampere: Tampereen yliopistopaino., 171-194.
- Smeds, R. (2003). *Helmi: Holistic Development of e-Learning and Business Models*. *Learn Periodical* 1/2003. Helsinki.

Development plan for education and university research for the period 1999-2004

<http://www.minedu.fi/julkaisut/KESU2004/eng/engKESU.html>

Education, Training and Research in the Information Society; A National Strategy for 2000-2004

<http://www.minedu.fi/julkaisut/information/englishU/index.html>

Information Strategy for Education and Research 2000-2004; Implementation Plan

[http://www.minedu.fi/minedu/education/inform\\_strategy\\_implementation\\_plan.pdf](http://www.minedu.fi/minedu/education/inform_strategy_implementation_plan.pdf)

The Strategy of the Finnish Virtual University

<http://www.virtuaaliyliopisto.fi/arkisto/FVU-strategy.pdf>

---

## Author Information

### Ms. Anna-Kaarina Kairamo

Helsinki University of Technology  
Teaching and Learning Development  
P.O.Box 8000  
FIN-02015 HUT  
Finland

anna-kaarina.kairamo@hut.fi  
www.hut.fi

## Grey Literature in an Open Context: From Certainty to New Challenges

C. Stock and J. Schöpfel  
INIST-CNRS

### Abstract

1997 grey literature was defined at Luxembourg as “that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers”. Distinction between commercial (white), non-commercial (grey) and other (dirty) literature seemed easy. Six years later, the landscape of scientific and technical information has radically changed. What is the place of grey literature between continuous electronic resources, open archives and new economic models of scientific literature? Is there any place left? The authors describe the evolution and relative decline of traditional grey literature from a French perspective and present some actual trends in the processing of grey literature by INIST-CNRS.

### A fairy tale of grey

Once upon a time, in a land called GreyNet ruled a mighty King whose name was Dominic. His TransAtlantic Kingdom reached to both sides of the Ocean, and its borderlines were well defined.

Outside was white literature, were serials and books easy to find, heavy price increases each year, commercial dissemination by publishers and vendors, the “dictate of the marketplace” and impact factors.

Inside were reports, conference proceedings, theses and other scientific and technical material often hardly to find. Inside were learned societies, research organizations and libraries, universities and corporate R&D, low-budget projects, rather nice and idealistic people sharing “common values about grey literature and (...) committed to preserving it” (Gelfand 1996).

Once upon a time, people knew more or less what grey literature was about. Because it was difficult to find, some created directories, reports or databases in order to facilitate access to these non-conventional, fugitive, and sometimes ephemeral publications.

In this Golden Age of Grey, studies stressed the importance and quality of this literature, of its strategic relevance for scientific and technological advances in all kind of subjects (see for example some recent articles: Conn et al. 2003 and Weaver et al. 2002 for the biomedical sector, Ramalho Correia & De Castro Neto 2002 for mathematics and agriculture, and Denda 2002 for women’s studies).

This was the era of large-scale and political projects. On the European level, EAGLE was created to facilitate the visibility of and the access to grey documents of the member states (see Wood & Smith 1993, Lankenau 1997). In France, the government decided 1990 the financial support of a national experiment for the centralized collection, description and dissemination of public reports that was conducted by a special committee under the responsibility of the Prime Minister (see CCDA 1995).

At INIST, we organised the acquisition, dissemination and archiving of French grey literature, building a unique collection of 300,000+ scientific reports, conference proceedings and dissertations in Sciences and Humanities, and most of them grey. As a member of EAGLE, we contributed to the SIGLE database (10% of the records are from INIST); as the French national ISRN agency, we tried to implement the international standard for report production and identification (see Schöpfel 2000).

## Grey paradise lost

All this seems long ago. Some years later, we doubt if this Golden Age of Grey had really existed. Was it a chimera, as ephemeral as its object?

The French report project completely disappeared, as did the national network for archiving and dissemination of French theses, without being substituted by a workable alternative. From 1998 to 2003, the acquisition of reports, proceedings and dissertations by INIST decreased by 50%, and the French ISRN agency is the only national agency left, without any significant support from the former international agency or the national standards organisation. Results of years of continuous and steady work are at risk of disintegration and decay.

At the same time, a sort of soft or gentle implosion threatens EAGLE, as its most important members are inclined to limit their engagement and cast doubts on the underlying economic model.

Even King Dominic disappeared, leaving his TransAtlantic Kingdom empty.

So what has happened? Why have we lost our grey paradise? What angel hunted us from our favourite playground?

In fact, there happened to be two angels, one called the worldwide web and the other, liberal politics.

### *The web*

Opening the 1<sup>st</sup> Plenary Session of the 2<sup>nd</sup> International Conference on Grey Literature in Washington, D.C., Julia Gelfand (1996) defined the web as the “new classical grey literature” and spoke of a dream that come true, empowering authors, increasing the role of libraries and serving the public good. But she already anticipated problems, the need of powerful research tools, of archiving, of standards, the risk of a growing flow of finally useless and non-relevant information, and the concurrence by commercial e-publishing.

Others shared this view. For instance, Judy McDermott (1996) from the Library of Congress stated in her keynote address at the same conference, “Just because information can be produced and easily distributed, it doesn’t necessarily mean it should have been” and went on, “Conversely, just because material is grey or hard to acquire doesn’t necessarily mean it’s worth the effort”.

This was prophetic; Julia Gelfand’s dream partly turned out to be a nightmare. If the document is on the web, why bother about it? Why invest into printing, supplying, archiving, and cataloguing? Putting the document on the Web is enough. Why continue collecting, recording, archiving non-conventional literature? Google, Yahoo or Altavista will do it for you. - How many stakeholders, once they heard of HTML and learned how to use a search engine, learned also to use this argument with and against information specialists and librarians?

### *Politics*

This was not without impact on public funding. Actually, most of our countries are in a difficult economic and financial situation. Governments redefine public priorities, cutting down public funding, rationalizing and setting for-profit goals to the public sector. Public research and Higher Education sectors are undergoing profound and rapid changes where “marginal” or “fugitive” objects and non-for-profit projects are hard to legitimate.

In France, funding of public research decreases, and the large public research organisations are forced to reconsider their functioning and objectives.

In this context, public funding of production and dissemination of non-conventional scientific documentation shifts from large-scale centralized projects to low budget decentralized networks and initiatives.

The coordination of dissemination of public reports exists no longer. Support for ISRN is vanishing, the national collection and online database of dissertations disappeared, the development of an international cooperation for grey literature is no longer a political but rather a sort of “private” objective of the participating organisations. But what if these institutions - universities for instance - don’t have enough money and have to face other priorities and urgencies? There is no powerful lobby to defend the “grey cause”.

### **Grey literature in post-modern age**

In this situation, King Dominic comes back and calls the grey citizens for a new round table, a new mission. But what is left from the once mighty Grey Kingdom? And where should we go?

Former certitudes about what is grey disappeared. Some continue to define grey literature in the traditional way, evaluating the part of grey documents at 10-15% of overall scientific production (see Leung 2002 or Weaver et al. 2002). Others describe grey literature as “materials not identifiable through a traditional index or database” and estimate the part of grey or “quasi-grey” production at 60+%, including for instance serials with limited geographical dissemination (see McKimmie & Szurmak 2002). Here, grey is not the document but the way to put a question and to search on the Web, especially the deep or invisible Web through special portals or search engines (see Parsons 2002). So, what is really grey?

With the certitudes disappeared standards, leaving a space of heterogeneity and diversity. Actually, there is nothing in the grey landscape that could be compared to the commercial Digital Object Identifier. Most projects that apply the emerging norm of metadata (Dublin Core) to grey objects are limited, often national or infra-national, or for only one special type of document.

Absence of certitudes, norms, and homogeneity - the Grey Kingdom entered in a post-modern period, in an “era without a dominant ideology but with a pluralism of style” where eclecticism and pragmatism are the new keywords (see van Raaij 1993).

In this post-modern era, instead of telling grey fairy tales or complaining for the past we need some step-stones that could be of help to progress on our way. In the following we’ll propose five subjects that may serve as such step-stones. Today there are more questions than answers, but you have to put the right questions if you want to progress.

### ***A new (or old) definition of grey literature?***

Ten years ago, Wood & Smith (1993) already had to cope with problems how to define grey literature. Instead of providing a clear-cut definition, they described it by means of several categories such as nature of material, way of acquisition and distribution, print number, format, standards of editing and production, publicity, bibliographic control, availability in libraries, informational value and promotion by producing organisations.

Mackenzie Owens (n.d.) states that “Grey does not imply any qualification, it is merely a characterisation of the distribution mode”.

In spite of a lot of conceptual work, obviously we can’t give a more specific answer to what is grey than ten years ago. The limits between continuing resources (e-serials, databases) and

monographs are moving. So then why not translate Wood & Smith's pragmatic approach into the reality of 2003? Instead of elaborating another (and probably useless) theory, another list of what is grey, why not re-evaluate the underlying categories of information production and dissemination, in order to apply them to those documents that are not submitted to commercial scientific edition?

Our proposition is to move from the traditional definition of grey literature in terms of document type and content (theses, conference proceedings etc.) to a multi-dimensional approach to the non-commercial part of scientific production, based on the distribution and access mode.

### *Economic model*

Collection and dissemination of grey literature never have been totally free. Someone has to pay, anyway. Even if access to grey literature could be described as different to traditional "book-selling channels", we must admit a specific commercial character to this activity. Even if the acquisition of dissertations or reports may seem to be free for libraries (of course, that's wrong), dissemination clearly is not, and most of the time the end user has to pay for the supply of the requested document. Often, this price is significantly higher than for information in commercial publications.

For instance, the price of a thesis goes from 50 to 100 euros or more; for a scientific report, the end user has to pay between 15 and 250 euros. This price is not linked to the quality or value of information nor to the price of the document itself but to the rarity of this information, to the relative small number of requests, to the monopolies of archiving and to the number of pages.

The underlying economic model remains uncertain. Let's take the preprints. Most preprints such as in Paul Ginsparg's ArXiv are free, but others are not (for instance, Springer's "Online First" service). So, the same product - for instance, an article in Mathematics that will be published by Springer - may at the same time be free of charge and high-priced? Is this a practicable economic model in the era of Web and open archives?

Our proposition is to analyse the production and dissemination of non-commercial documents in terms of investment, direct and indirect costs, and not only in terms of prices, in order to contribute to one or more economic models for non-conventional literature.

### *Archiving - but what, and how, and who?*

The grey community always shared the conviction that grey literature should be archived because of its fugitive and ephemeral nature. But three aspects need to be considered.

First of all, it is difficult to decide what should be archived without considering the risk of a growing flow of finally useless and non-relevant information (see Gelfand 1996). The same document may exist in more than one version, some of them or all grey. Should all versions be archived? Should we archive work in progress? Why should we do so?

Second, the shift to electronic resources doesn't facilitate the task. Because of the rapid evolution of hard- and software, of supports and formats, and without accepted standards, archiving is discussed rather in terms of 5-10 years than of decennials or centuries. Ironically, we assist to a generalisation of ephemeral information. But for already fugitive documents, this generalisation carries the risk of disappearance.

Archiving not only needs important investments but also a continuity of service. Third question: who should bear the costs of archiving non-conventional literature? There are no commercial publishers to do it. Should it be the producer of information (HE, public and corporate R&D, the

researcher-author himself)? Or should it be some central operators (national libraries or information centres)? We already described the difficult financial situation of the public sector. Most often, priorities are given to commercial publications. Result: in the emerging world of web-based archives, conference proceedings, reports and dissertations today are often stored on ephemeral, fugitive and disappearing servers. Tomorrow, what will be left of all this?

Our proposition is to rethink the question of grey archives in the emerging environment of open archives (see Grüttemeier & Mahon 2003).

### ***Quality issues and value chain***

The problem of the quality and interest of non-conventional literature is well known. The last GL conference discussed the concept of “dirty grey”. But how distinguish grey from dirty grey or black (= web trash)? Can this question be discussed only in terms of content quality?

Roosendaal et al. (2002) described the transformation of the traditional value chain of scientific information into an emerging model with a clear shift in added value to the education and research organizations. They predicted the shift to take place for the serials in the next five to ten years.

For grey literature, the key issue of certification should be addressed differently. Roosendaal’s distinction between an “author system” where the text is submitted to a refereed publication medium, and a “reader system” where the text is accepted or not in an open system seems more complex for grey documents. Frequently, the same “key actors” that requested or supported their publication provide their acceptance or approval (research organisations, universities, corporate R&D etc.).

However, it is not always clear if these institutions act as “collective non-commercial referees” or as “collective readers”. Often, they do both. Should we, then, speak of a “referee-reader-system”? Or more exactly, of a “two-level-reader-system” where first approval is provided by collective institution (submission) and further acceptance is requested from individual reader (for instance per auto-archiving), both of them being part of the same scientific community?

Additionally, Mackenzie Owen (n.d.) proposes a quality control for grey literature provided by “innovative technologies” such as rating systems or electronic refereeing.

Our proposition is to shift the quality discussion from content evaluation to Roosendaal’s value chain approach, introducing the concept of approval and certification.

### ***Copyright issues***

Often, grey documents are considered free (or outside) from copyright restrictions. However, reality is more complex, for two reasons. First of all, all these documents bear individual or moral authorship, with specific intellectual rights protected by national law, even if they are not always clearly defined: who owns for instance the intellectual rights for the final report of a public funded research project?

On the other hand, more and more funding bodies - public as well as corporate - are inclined to protect the intellectual content of grey documents, not because of national law but in order to protect the results of scientific research itself. In this context, it would be interesting to analyse the relationship of available grey material and the increasing number of patents. For example, at INIST we are confronted with a relative decline of research reports published by major research institutions and the corporate sector because reports and even university material are classified, with more or less restrictive distribution.



It seems difficult to have a clear vision of the legal aspects of grey literature, especially during the ongoing transposition of the European copyright directive (see FIPR 2003). Our last proposition is to develop a “legal analyse” of grey production in terms of “accessibility”, in order to avoid obscure and erroneous action.

### Searching for new horizons

After all, is there any place left for our grey community? Or has the time come to take a definite leave from the TransAtlantic Kingdom and to move King Dominic to Mme Tussaud’s cabinet?

We assist to a rapid evolution of the landscape of scientific and technological information. But in the era of concentration of commercial edition and big deals, we should be aware that crucial parts of information rest marginal, excluded from gateways and consortial purchasing and difficult to access through metasearch engines or virtual libraries.

We have to redefine our action of making available and preserving hidden or unconventional information. Grey topics should be discussed and re-evaluated in terms of investment and economic models, of quality and certification, of rights and protection, and of accessibility.

What are our affinities with emerging projects such as SPARC and OAI? Some goals of SPARC projects are quite similar to grey projects (see <http://www.arl.org/sparc>). SPARC and BiomedCentral develop new economic models that challenge the for-profit STM publishers. The Open Access Initiative tries to offer an answer to some of our own grey questions, by supporting auto-archiving in a standard format and e-publishing in OA publications.

The French National Research Centre CNRS co-organized the first French conference on alternative e-publishing models in January 2003 at Paris (see <http://www.inist.fr/openaccess>, Grüttemeier & Mahon 2003 or Battisti 2003) and signed, together with the German Max-Planck Gesellschaft and other important European research organisations, the Berlin Declaration on open access to knowledge in the sciences and humanities. This means a fundamental paradigm shift of funding, organisation and evaluation of scientific publication that will change the market of scientific information and the role of libraries, in a way predicted by Mackenzie Owen (n.d.) and Roosendaal et al. (2002).

We are convinced that in the emerging Empire of Virtual Information, our former Grey Kingdom could become a sort of unconventional region or structure, with a specific and different contribution to this new evolution. “In a world in which free trade and instantaneous communication have eliminated many of the barriers to information flow, grey literature is gaining greater importance as a source of information for much of the world’s population. It is an indispensable resource for an informed and enlightened public and will undoubtedly continue to serve as a necessary supplement to journal literature well into the future” (Weintraub 2000).

We need to rethink and reshape our “grey approach” in the context of open archives and free access to scientific results. Learning from model projects may be of help. For instance, the gateway of the French Mathematics libraries “MathDoc” (<http://mathdoc.emath.fr>) created an index of 4,000+ grey documents (see Lafosse 2003, Sureau & Teissier 2003). The Open Library for Economics RePEc (see Krichel 2000) contains 80,000+ working papers. The New York Academy of Medicine (<http://www.nyam.org/library/greylit>) supports an important directory of grey literature in the medical sector. The extending German “Virtual Subject Libraries” include grey resources (see Rosemann 2003), as well as the recent INIST portals in Social Sciences and Humanities and in Life Sciences. And last not least, the web-based American GrayLIT Network gives access to over 100,000 recent scientific and technical reports by the US Departments of Energy and Defense, the Environment Protection Agency and the NASA (see Warnick n.d.).

## Projects@INIST in a changing context

The impact of this changing environment on INIST's services and projects has been described elsewhere (see Schöpfel 2003). In dealing with grey literature in paper form INIST has often been confronted with difficulties in collecting documents mainly because of an insufficient number of available copies. The copies were stored on our shelves and referenced in several databases such as PASCAL or SIGLE. The user had to order a photocopy at INIST.

With electronic grey literature, copies are no longer a problem. Many producers of grey literature make their documents available on the Internet: either the reference, with or without abstract, or the full text. Very often the documents are found on the institutional server under insignificant headings such as "list of publications of the institute".

INIST started several initiatives attempting to concentrate dispersed references in a unique database, to incite authors to publish electronically and to deposit the documents in reliable repositories, and to allow users an easy access to the full text and free of charge.

In the second part of our communication, we'll describe three projects that change the traditional acquisition and dissemination of grey literature, and in which INIST participates or intends to participate.

The three projects are:

**MemSIC:** collection and references of theses in information science, documentation and communication.

**TEL:** a French server for dissertations online.

**LARA:** localisation of and access to reports - a project still in its beginnings.

## The MemSIC project

For several years now INIST has been collecting "thesis" ("mémoires" in French) from students in information science and documentation ("SIC" in French) at a Masters degree level. The documents were received in paper form, catalogued and referenced in the PASCAL and SIGLE databases, and made available to users, locally and by photocopy.

Since 2000, INIST developed together with the university of Nancy a web-based archive for this type of documents called "MemIST" ("IST" = scientific and technical information). At present, we are working together with the new CNRS Centre for Direct Scientific Communication at Lyon (Centre pour la Communication Scientifique Directe or CCSD, see Charnay 2003), created in 2000 by Franck Laloë, a researcher in Physics collaborating with Paul Ginsparg's ArXiv (see <http://www.ccsd.cnrs.fr/>). The objective of this project is to load the prototype on a CCSD server in order to set up a specialized repository named "MemSIC" that will give access to the full text (<http://MemSIC.ccsd.cnrs.fr/>).

## Deposit

Students who want to submit their thesis must be registered as a user. Access to the deposit function is password protected. A worksheet allows the student to fill in the metadata. Fields are based on the DC metadata set, but some fields have been added for this specific type of document such as the type of the diploma. In a second time these metadata will be integrated into other databases (PASCAL, SIGLE).

In the next step of the deposit procedure the student chooses among 22 subject categories, then he selects the document format. The document formats allowed at present are DOC, RTF, PDF, PS and HTML. Each document is immediately transformed - if necessary - into the PDF format for dissemination.

### ***Copyright and validation issues***

By activating the submit button, each author agrees to disseminate his work via this archive, transferring de facto his intellectual rights to the CCSD.

The head of the programme adds institutional authorization: before making each deposited document available for the public, it must be validated by the head of the programme. It's a simple procedure by exchange of e-mail with the administrator of the archive who has the final decision to make the thesis accessible.

This institutional validation guarantees the scientific approval of the document's content quality.

### ***Coverage, search tools and environment***

MemSIC is open to university programmes in information and communication sciences in France and in French speaking countries, on a Master's degree level.

The interface contains the usual search criteria (title, author, abstract, keywords, subject category, institution, diploma type...) and a simplified keyword search. MemSIC allows browsing by author and subject category and full-text search in the document. By subscribing to the archive, the end user is alerted on each new deposit.

MemSIC is embedded in a larger environment of open archives for information and communication sciences with a repository for articles, communications and preprints ("ArchiveSIC" <http://archiveSIC.ccsd.cnrs.fr/>) and for doctoral dissertations ("TheSIC"). A common search interface "SIC" for these 3 archives will be added soon.

### ***Why this network?***

Few websites have stable URL'S for their lists of publications, and the guarantee of perennial access to the document is low. The readability of the format will arise as another problem in a few years. Several private initiatives originally created to allow auto-archiving of theses and dissertations without institutional control have been abandoned since.

INIST had to choose either to install an open repository or to work with a reliable institution in order to assure a continuity and stability of document access.

Our partner, CCSD, provides an excellent technical environment. A customized version of the e-prints software is used for the submission of documents. It proposes facilities for the deposit, metadata storage and exchange (use of protocols like the OAI-PMH), powerful search software, user-friendly interfaces, personal workspace, alert function and statistics. Last but not least: CCSD has a mission for long-term archiving, thus providing a "guarantee" for the migration of formats and for permanent access.

The main role of INIST will be to promote the project, to "collect" the documents by contact of the heads of the study programmes and to administrate and develop the database. In addition,

INIST will load the metadata of the grey documents into its own databases and portals, increasing their visibility and dissemination to the scientific community.

### The TEL project

MemSIC widely uses features and elements already developed for “TEL”, another CCSD archive, namely the metadata set and the deposit procedure.

CCSD started as the French mirror site for Paul Ginsparg’s ArXiv (publications/articles in physics). Soon it developed towards a general site for open archives, with extensions in the scientific fields as well as in document types.

Two of the developments were the setting up a repository for theses and dissertations “TEL” and hosting a repository for articles in information science and documentation (“ArchiveSIC”).

Logically, TEL (“Theses on-line”, <http://tel.ccsd.cnrs.fr/>) started with dissertations in physics, mathematics and computer science, disciplines used to open archives. Today documents are grouped in 9 scientific fields, with several sub-categories. The majority still comes from physics as shown in the list below (data as of November 17th, 2003):

Physics	761
Mathematics	174
Computer Science	97
Engineering Sciences	79
Sciences of the Universe	62
Biology and Medicine	39
Chemical Sciences	29
Humanities, Social Sciences	20
Other areas	14

A presentation of documents by university of origin shows that the site is open to foreign documents. We find dissertations from Great Britain, Germany, Spain, Switzerland, etc.

At present several public or private initiatives that already give access to dissertations online intend to migrate to TEL or to duplicate and link their sites with TEL. One example: an important consortium of 10 engineering and commercial schools of the Paris region “ParisTech” (= Paris Institute of Technology) who created a repository for dissertations called PASTEL actually proposes to reference these documents also in TEL.

After the disappearance of the former national network for archiving and dissemination of French doctoral dissertations, INIST decided to support the CCSD initiative rather than to create an own archive. In this context, the role of INIST is twofold: incite other well known institutes giving access to dissertations in full text to deposit them in TEL, and reference these documents through its own products and services (portals, databases).

As mentioned before, dissertations in information science, as a subset of TEL, will also be searchable through the SIC interface.

The advantage of projects like MemSIC or TEL is that they are more likely to succeed than others: students are young, willing to learn and accustomed to produce electronic documents. In addition, institutional pressure (“no deposit, no diploma”) facilitates the input to the open archive.

## The LARA project

The third project, LARA (= “localization of and access to reports”), is only in its beginning and concerns scientific and technical reports. In this case we intend to create a bibliographic database for French reports available online or in paper form by using the OAI protocol for metadata harvesting.

This means that INIST limits its contribution to identifying the report producers, to set up the metadata harvesting procedures and to provide a server. In some cases, the documents will be archived on an INIST server, in their original electronic format, converted to a standard format or digitised by INIST. In other cases, the document supply or the access online to the full text will be left to the publishing organization or the document producer. Of course, agreements with these organizations will be necessary.

Data collected by metadata harvesting will be linked with our own existing report catalogue and will be searchable through a web server.

The model for LARA is the UK MAGiC project developed by the British Library and Cranfield University and focussed on reports in the aeronautics sector (see Needham 2002). INIST is in contact with the Cranfield project team. Maybe that this will give place to a transnational collaboration in the field of electronic reports.

Up to date a prototype has been created and metadata harvesting has been tested. Finally, the success of this project will depend on the willingness of public and corporate report producers to cooperate (several have shown their interest), and on the progress for implementing the OAI-MH protocol in France.

## Open = free?

In the open archive context information is virtually free for the end user, except costs for downloading and printing of files. Nevertheless, there are costs involved for setting up repositories, developing interfaces, updates of software, metadata harvesting, referencing in portals and databases, promotion, contact with document producers, preparation of agreements, providing technical assistance, etc.

Obviously, these costs can only be supported by a public structure that guarantees a continuity of qualified service and action in a sector without any commercial interests or returns. In the actual economic context we described above, INIST tries to rationalize and optimise the technical and organisational investment, sharing costs and using existing network resources rather than create a new centralized structure. The future will show if this option is practicable and effective, and if this new approach will allow a long-term and secure access to ephemeral and fugitive, non-conventional scientific literature.

## Bibliography

Battisti M. (2003): “Libre accès à l’information scientifique et technique : état de l’art et perspectives”. *Documentaliste - Sciences de l’Information*, Vol.40, No.1, pp.37-45.

*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* (2003).

<http://www.zim.mpg.de/openaccess> (visited 2003-10-25).

CCDA Commission de Coordination de la Documentation Administrative (1995) : *Synthèse de l’expérience littérature grise des administrations 1990-1994*. Paris : Premier Ministre, Secrétariat général du gouvernement.



- Charnay D. (2003): "The Centre for Direct Scientific Communication". In Grüttemeier H. & Mahon B. (2003), pp. 133-137.
- Conn V.S., Valentine J.C., Cooper, H.M., Rantz, M.J. (2003) : "Grey Literature in Meta-Analyses". *Nursing Research*, Vol.52, No.4, pp.256-261.
- Denda K. (2002): "Fugitive Literature in the Cross Hairs: An Examination of Bibliographic Control and Access". *Collection Management*, Vol.27, No.2, pp.75-86.
- Farace D. (1996): *GL'95 proceedings. 2th International Conference on Grey Literature. Grey Exploitations in the 21<sup>st</sup> Century. Washington D.C., 2-3 November 1995*. Amsterdam: TransAtlantic/GreyNet Grey Literature Network Service.
- Farace D. (2000): *GL'99 proceedings. 4th International Conference on Grey Literature. New Frontiers in Grey Literature. Washington D.C., 4-5 October 1999*. Amsterdam: TransAtlantic/GreyNet Grey Literature Network Service.
- FIPR Foundation for Information Policy Research (2003), *Implementing the UE Copyright Directive*, pp. 55-69, available at: <http://www.fipr.org/copyright/guide>. (visited 2003-10-09).
- Gelfand J. (1996): "Grey Literature in New Packages: Implications from the Transition to Electronic Publishing and Information Seeking". In: Farace D. (1996), pp. 18-32.
- Grüttemeier H. & Mahon B. (2003): *Open Access to Scientific and Technical Information*. Amsterdam: IOS Press.
- Krichel T. (2000): "RePEc, an Open Library for Economics". <http://openlib.org/home/krichel/papers/salisbury.html> (visited 2003-11-12).
- Lafosse Y. (2003): "La Cellule MathDoc". <http://www.inist.fr/openaccess> (visited 2003-10-09).
- Leung Y.C. (2002): "Always Yes, Never No". *Journal of Interlibrary Loan, Document Delivery & Information Supply*, Vol.12, No.3, pp.13-27.
- Mackenzie Owen J. (n.d.): "The Expanding Horizon of Grey Literature". <http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/glpaper.htm> (visited 2003-10-21).
- McDermott J.C. (1996): "Defining Grey Literature for the 21<sup>st</sup> Century". In: Farace D. (1996), pp. 9-17.
- McKimmie T., Szurmak J. (2002) : « Beyond Grey Literature : How Grey Questions Can Drive Research". *Journal of Agricultural & Food Information*, Vol.4, No.2, pp.71-79.
- Lankenau, I. (1997): "Grise mais disponible: le système d'information sur la littérature grise en Europe (SIGLE)". *Revue d'Information Scientifique et Technique*, Vol.7, No.2, pp.101-109.
- Needham P. (2002): "MAGiC: Shining a New Light on a Grey Area". *Serials*, Vol.15, No.3, pp.201-206.
- Parsons M. (2002): "Free Full-Text Resources on the Web for Small Libraries". *1<sup>st</sup> Northwest Interlibrary Loan/Resource Sharing Conference*. Portland, Oregon, September 20<sup>th</sup>, 2002.
- Raaij W.F. van (1993): "Postmodern consumption". *Journal of Economic Psychology*, Vol.14, No.3, pp. 541-563.
- Ramalho Correia A.M., De Castro Neto M. (2002): "The Role of ePrint Archives in the Access to, and Dissemination of, Scientific Grey Literature: A Case Study by the National Library of Portugal". *Journal of Information Science*, Vol.28, No.3, pp.231-241.
- Roosendaal H.E., Geurts P.A.Th.M., van der Vet P.E. (2002): "Integration of Information for Research and Education: Changes in the Value Chain?" *Serials*, Vol.15, No.1, pp.51-56.
- Rosemann U. (2003): "Die Arbeitsgruppe der Informationsverbände und die Virtuelle Fachbibliotheken: Beginn einer wunderbaren Zusammenarbeit?" *Zeitschrift für Bibliothekswesen und Bibliographie*, Vol.50, No.1, pp.13-18.
- Schöpfel J. (2000): "INIST: Tracking Grey Literature in a Changing Environment". In: Farace D. (2000), pp. 199-202.
- Schöpfel J. (2003): "INIST-CNRS in France: 'A Model of Efficiency'". *Interlending & Document Supply*, Vol.31, No.2, pp.94-103.
- Sureau G., Teissier B. (2003): "Le Réseau national des bibliothèques de mathématiques". *Bulletin des Bibliothèques Françaises*, Vol.48, No.2, pp.56-58.
- Warnick W. (n.d.): "Tailoring Access to the Source: Preprints, Grey Literature and Journal Articles". <http://www.nature.com/nature/debates/e-access/Articles.warnick.htm> (visited 2003-10-01).

- Weaver N., Williams J.L., Weightman A.L., Kitcher H.N., Temple J.M.F., Jones P., Palmer S.: "Taking STOX: Developing a Cross-Disciplinary Methodology for Systematic Reviews of Research on the Built Environment and the Health of the Public". *Journal of Epidemiology and Community Health*, Vol.56, No.1, pp.48-55.
- Weintraub I. (2000): "The Role of Grey Literature in the Sciences".  
<http://library.brooklyn.cuny.edu/access/greyliter.htm> (visited 2003-10-21).
- Wood, D.N., Smith, A.W. (1993): "SIGLE: A Model for International Co-operation". *Interlending & Document Supply*, Vol.21, No.1, pp.18-22.

---

## Author Information

### Christiane Stock

INIST-CNRS, Institut de l'Information Scientifique et Technique  
Centre National de Recherche Scientifique  
2, allée du Parc de Brabois  
54514 Vandoeuvre-Les-Nancy Cedex  
France

[schopfel@inist.fr](mailto:schopfel@inist.fr)  
[christiane.stock@inist.fr](mailto:christiane.stock@inist.fr)  
[www.inist.fr](http://www.inist.fr)  
[www.cnrs.fr](http://www.cnrs.fr)

# CERN Document Server: Document Management System for Grey Literature in Networked Environment

Martin Vesely, Thomas Baron, Jean-Yves Le Meur and Tibor Simko  
CERN

## Abstract:

In this paper we present a technology for networked information services, developed at the CERN Document Server (CDS) research group, called the CERN Document Server Software (CDSware).

Standardization of networked information services in the field of grey literature has recently become a subject of an intensive research in the digital library community. The current state-of-the-art in this area effectively allows to provide various networked information services, such as information brokering or other value-added services upon distributed or federated data. This refers specifically to a variety of newly developed frameworks, such as the Open Archives initiative Protocol for Metadata Harvesting (OAI-PMH).

The CDSware technology offers a comprehensive solution for a document management of a large grey literature document repository, compliant with a variety of networking standards essential for a wide deployment of networked information services.

## 1 Introduction

The research in the field of digital libraries has recently focused on information interoperability models, the integration of distributed and federated data and networked information management. One of the key features of the networked information is that data, information and knowledge can be gathered, processed, stored and maintained separately from the information services offered by information mediators or brokers. In the scope of scholarly communication the models of networked information involve predominantly distributed and federated data processing, built on top of various widely deployed internet technologies, transfer protocols, their extensions and finally also comprehensive technological frameworks such as the Web Services.

Within this perspective, the Open Archives Initiative developed a Protocol for Metadata Harvesting [OAI-PMH]. The OAI-PMH was set up in 1999 [Van de Sompel, 2000] in order to filter out information heterogeneities that prevented an efficient cooperation between various e-print archives and other grey literature repositories in the networked environment. The interoperability is achieved by sharing the metadata format schemata used by all parties involved. One of the important goals of this protocol is to allow federated archives to harvest references to relevant documents that can then be made available for the archive users.

The CERN Document Server (CDS) group has been active in research focusing on interoperability of digital document storage and retrieval systems, particularly promoting the WWW and related technologies in the digital library community, including the OAI-PMH. Within the last 5 years the digital library research at CERN focused on (i) linking strategies to digital library resources with emphasize on the scientific literature [Vigen, 1999], (ii) data integration from heterogeneous data sources [Vesely, 2000], (iii) specification of the protocol for metadata harvesting [Vesely, 2002] and (iv) automated indexing of scientific documents [Dallman, 1999] [Raez, 2002].

## 2 Grey literature management

Until recently, the management of grey literature collections and mediation of scientific information has been performed predominantly by specialized *disciplinary repositories* based mainly on a centralized model. One of the pioneering repositories of this type was the ArXiv.org



repository, specialized in the domain of high-energy physics, later also in other scientific disciplines. Since that time, *institutional repositories* have been gaining more and more importance [Lynch, 2003]. In the scope of this framework, document management in institutional repositories has been targeted by the CDS research group at CERN that has led to a construction of a freely available modular on-line document system software [CDSware] presented in this paper.

The current state of the art in grey literature management and the currently available technologies allow to provide syntactic and structural interoperability allowed in the first place by the XML related specifications. On the other hand, the semantic interoperability has not yet been scientifically resolved. The current trend goes towards attempts to address these issues, mainly by initiatives such as Semantic Web [SemanticWeb] or Semantic Grid [SemanticGrid]. As an example of a promising standardization effort in the semantic interoperability of digital libraries, the Dublin Core metadata format can be cited [DublinCore] that is now becoming a de-facto standard in the new open global network beyond the scope of the digital library field where it was first proposed. For example, it now appears as a convenient format for metadata encapsulation in the scope of RDF/XML.

Following the digital library research, we attempt to apply the resulting scientific achievements in implementations of concrete systems that allow users to benefit from the most recent scientific results and the leading edge technology. Since the year 1996 when the CDS section has been established at CERN we have been supporting the local digital library system with innovative technologies, and most recently, we have been working on integrating these technologies to a coherent document management system.

### 3 Modules

The CDSware technology aims at a management of a middle-sized repository of documents, representing a repository size up to  $10^5$ - $10^6$  records. CDSware consists of several independent modules with precisely defined functionality (see Figure 1). The general criterion for module names is to use the ``Bib" prefix to denote modules that work more with the bibliographic data, and the ``Web" prefix to denote modules that work more with the Web interface. (The difference is of course blurred in some cases, as in the case of search engine module called WebSearch that has got a web interface but searches bibliographic data.).

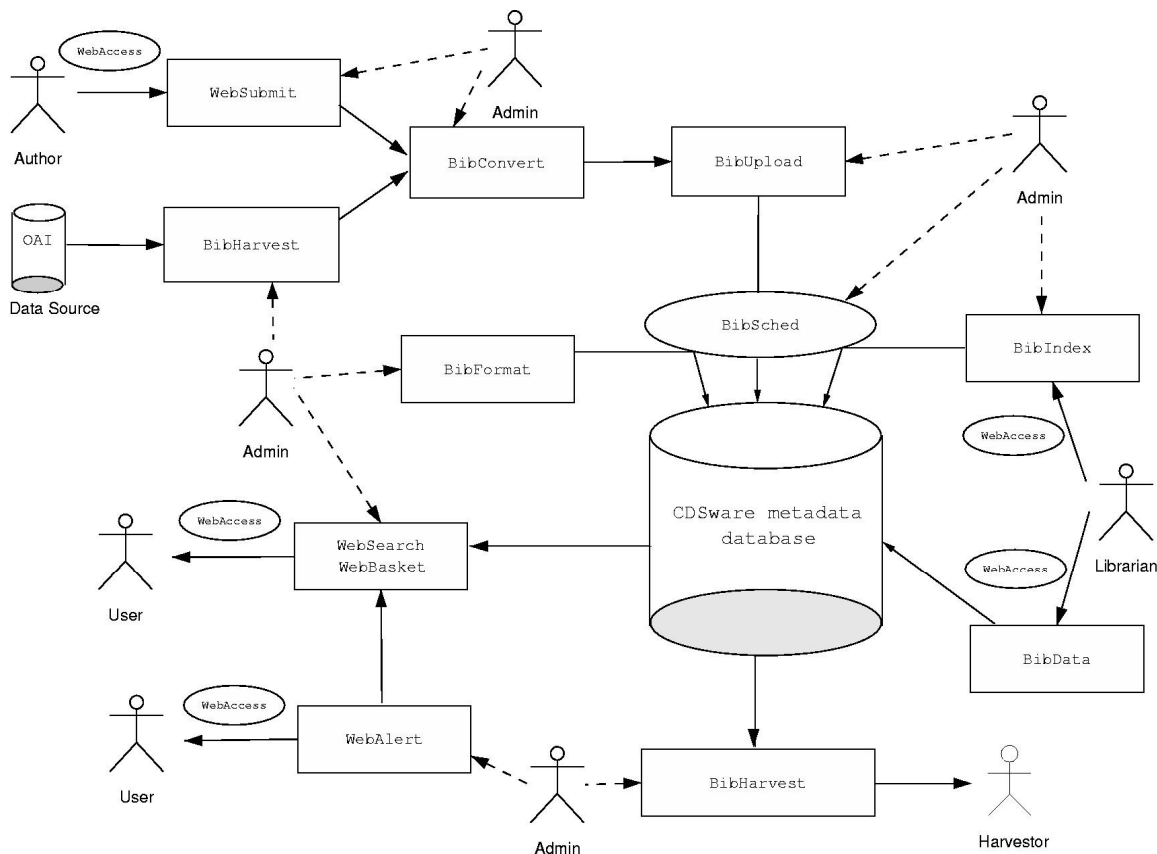


Figure 1. Overview of CDSware modules

**3.1 BibHarvest** represents the OAI-PMH compatible harvester allowing the repository to gather metadata from fellow OAI-compliant repositories and the OAI-PMH repository management. Repository is built directly on top of the database and disposes of an OAI repository manager that allows to perform the administrative tasks on the repository aside from the principal generic data administration module. The database can be partially or completely open for harvesting in the scope of the OAI-PMH protocol. In this case, all data is provided in raw form, where the semantics of individual tags is indicated uniquely by the MARC21 naming convention. This is particularly interesting for institutes that are specialized in cross-archive and cross-disciplinary services provision, as for example the ARC service provider [Zubair, 2001].

**3.2 WebSubmit** is a comprehensive submission system allowing authorized individuals (authors, secretaries and repository maintenance staff) to submit individual documents into the system. The submission system disposes of a flow-control mechanism that assures the data approval by authorized units. In total there are several different exploitable submission schemas at a disposal, including an automated full text document conversion from various textual and image formats. This module also disposes of information extraction functionality, focusing on bibliographic entities such as references, authors, keywords or other implicit metadata.

**3.3 BibConvert** allows metadata conversion from any structured or semi-structured proprietary format into any other format, typically the MARCXML [MARCXML] that is natively used in CDSware. Nevertheless the input and output formats are fully configurable and have been tested on data importations from more than one hundred data sources. The power of this utility lies in the fact that no structural attributes of data source are presumed, but they are defined in an extensive data source configuration. Inevitably, this leads to a high complexity of the BibConvert configuration language. Most frequent configurations are provided with the CDSware distribution, such as a sample configuration from Qualified Dublin Core into the MARCXML.

In general the BibConvert configuration consists from the source data descriptions and target data descriptions. The processor then analyzes and parses the input data and creates the resulting data structure, similarly as the XSLT processor would do. Typically the BibConvert is aimed at usage for input data that do not dispose of an XML representation. The source data is required to be structured or semi-structured, (i.e. not expressed in natural language that is a subject of information extraction task) and its processing involves several steps including record separation and field extraction up to transformation of source field values and their formatting.

**3.4 BibFormat** presents a very powerful module to format the bibliographic metadata in numerous ways. This truly enables the separation of data content administration and formatting layout, effectively allowed by the XML technology. Formatting can either be carried out as a batch task or invoked on the fly. The formatter allows a dynamic creation of automatic links according to destination descriptions.

**3.5 BibSched** The bibliographic task scheduler is central unit of the system that allows all other modules to access the bibliographic database in a controlled manner, preventing sharing violation threats and assuring the coherent execution of the database update tasks. The module comes with an administrative interface that allows to monitor the task queue including various possibilities of a manual intervention, for example to re-schedule queued tasks, change the task order, etc.

**3.6 BibUpload** allows to load the new bibliographic data into the database. To effectuate this task the data must be a well-formed XML file that complies with the current metadata tag selection schema. Usually, the properly structured input files of BibUpload come from the BibConvert utility.

**3.7 BibIndex** module takes care of the indexation of metadata, references and full text files. Two kinds of indexes -- word and phrase index -- are being maintained. The user can define several logical indexes (e.g. author index, title index, etc.) and the correspondence of which physical MARC21 metadata tag goes into which logical field index. An index consists of two parts: (i) a forward index listing various words (or phrases) found in the given field, with the set of record identifiers where the given word can be found; and (ii) a reverse index listing record identifiers, with the set of words of the given record that go to the forward index. Such a two-part indexing technique allows one to rapidly update only those words that have changed in the input metadata record. The indexes were designed with the aim to provide fast user-response search times and are faster than native MySQL (full text) indexes.

**3.8 WebAccess** module is responsible for granting access to users for performing various actions within the system. A Role-Based Access Control (RBAC) technique is used, where users belong to several groups according to their role in the system. Each user group can be granted to perform certain actions depending on possible one more action arguments. WebAccess is presently used mainly for the administrative interface. There are basically two kinds of actions: (i) configuration of administrative modules and (ii) running administrative tasks.

**3.9 WebBasket** module enables the end user of the system to store the documents she is interested in in a personal basket or a personal shelf. The concept is similar to popular shopping carts. One user may own several baskets. A basket can be either private or public, allowing a simple document sharing mechanism within a group.

**3.10 WebAlert** module allows the end user to be alerted whenever a new document matching her personal criteria is inserted into the database. The criteria correspond to a typical user query as if it would be done via the search interface. For example, a user may want to get notified whenever a new document containing certain words, or of a certain subject, is inserted. A user may create several alerts with a daily, weekly, or a monthly frequency. The results of alert searches are either sent back to the user by email or can also be stored into her baskets.

**3.11 WebSearch** module handles user requests to search for a certain words or phrases in the database. Two types of searching can be performed: a word search or a phrase search. The system allows for complex boolean queries, regular expression searching, or a combined metadata, references and full text file searching in one go. Users have a possibility to browse for present index terms. If no direct match could have been found with the user-typed query pattern, the system proposes alternative matches as a search guidance. The search indexes were designed to provide fast response times for middle-sized data collections of up to  $10^6$  records. As an example, Table 1 presents typical search times on the production CERN Document Server for several simple word queries as well as a boolean one.

Search pattern	No. of results retrieved	Search time
Ellis	1,760	0.21 sec
Model	84,168	0.22 sec
Cern	219,530	0.22 sec
Of	426,848	0.23 sec
of cern model	6,907	0.44 sec
of of of of	426,848	0.53 sec

Table 1. Typical search engine performance at the CERN Document Server production installation <<http://cdsweb.cern.ch/>>. The total number of documents in stored the repository is about 680,000. The search times were obtained over several runs so that the results are speeded up by the database server cache. Searching for virgin words may be slower by a factor of 5. Note the fast performance of boolean queries due to a special set-intersection-driven words index design. The tests were performed on a web application server running on dedicated Dual Pentium III 1.0 GHz (Ultra 160 SCSI) with database server running on Dual Pentium III 1.4 GHz.

The metadata corpus is organized into metadata collections that are directly accessible through the browse function, similarly to the popular concept of Web Directories. Orthogonal views on the document corpus are enabled in the search interface via a concept of virtual collections: for example, a document may be classified both according to its type (e.g. preprint, book) and according to its Dewey decimal classification number. Such a flexible organization views allows for the creation of easy navigation schemata to the end users.

#### 4 Statistics

It may be interesting to mention typical usage statistics of the CERN Document Server. The server holds more than 680,000 documents of interest to people working in particle physics and related areas. The server contains many kinds of metadata: preprints, articles, books, journals, photographs, and more. The grey literature (preprints) counts of up to 300,000 documents with more than 250,000 full text files. There are currently 4 open OAi metadata sets covering a total of ca. 40,000 harvestable documents.

The server is used by a community of particle physicists at CERN and all around the world. The typical usage indicates that the server is dominantly used by non-CERN clients (some 70%). There are typically over 200,000 distinct users per year coming from many countries and around 10,00 distinct visitors per month. On a typical day (30<sup>th</sup> September 2003) we have seen 1,417 unique visitors doing 4,870 searches; 16 distinct OAi harvesters connected to the site during that day making a total of 275 OAi requests.

#### 5 Conclusions

In this paper we briefly described the current status of research in the field of digital libraries interoperability, mainly with respect to federated data processing such as permitted within the Open Archive Initiative framework. We then presented the CERN Document Server Software suite that is a free software package maintained by CERN providing an online digital library

solution for mid- to large- sized document repositories. The set of CDSware modules was described and the differences with other existing tools briefly mentioned. Finally we have mentioned a typical usage statistics illustrating the scope of grey literature and networked conditions of the CERN Document Server.

We pointed out that the current trend in grey literature management goes towards institutional repositories build upon the distributed and federated model. Further research is going on within the CDS collaboration in this area, focused particularly on (i) semantic interoperability using ontologies in distributed and federated data processing and (ii) ranking technologies when searching grey literature in OAI compliant repositories.

### References:

- [CDSware] CERN Document Server Software. <http://cdsware.cern.ch/>
- [Dallman, 1999] Dallman D. and Le Meur J.-Y. Automatic Keywording of High-Energy Physics. *4<sup>th</sup> Intl Conference on Grey Literature: New Frontiers in Grey Literature GL99, Washington D.C., 1999.*
- [DublinCore] Dublin Core Metadata Initiative, <http://www.dublincore.org/>
- [Lynch, 2003] Clifford A. Lynch. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL*, no. 226 (February 2003): 1-7.  
<<http://www.arl.org/newsltr/226/ir.html>>.
- [MARCXML] Marc XML metadata format <http://www.loc.gov/marc/marcxml.html>
- [OAI-PMH] Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH), <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [Raez, 2002] Raez A.M. Toward conceptual indexing using automatic assignment of descriptors. *In: 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems. Personalization Techniques in Electronic Publishing on the Web: Trends and Perspectives, Málaga, Spain, 28 May 2002.*
- [SemanticGrid] Semantic Grid Initiative. <http://www.semanticgrid.org/>
- [SemanticWeb] Semantic Web Initiative. <http://www.w3.org/2001/sw/>
- [Van de Sompel, 2000] Van de Sompel H. and Lagoze C. The Santa Fe Convention of the Open Archives Initiative. *D-LIB Magazine*, Volume 6, Number2, 2000.
- [Vesely, 2000] Vesely, M et al. Using Internet/Intranet Technologies in Library Automation. *Thesis*, University of Economics, Prague, Czech Republic, 2000.
- [Vesely, 2002] Vesely, M et al. Creating Open Digital Library Using XML: Implementation of the OAI-PMH Protocol at CERN. *Intl. Conference on Electronic Publishing, Elpub 2002*, Karlovy Vary, Czech Republic.
- [Vigen, 1999] Vigen J. et al. Link Managers for Grey Literature. *4<sup>th</sup> Intl Conference on Grey Literature: New Frontiers in Grey Literature GL99, Washington D.C., 1999.*
- [Zubair, 2001] Zubair M, Maly K, Liu X. Arc, the OAI service provider for Digital Library Federation. *D-Lib Magazine*, Volume 7, Number 4, 2001.

---

### Author Information

#### Martin Vesely

CERN Document Server  
The European Centre for Higher Energy Physics  
P.O. Box ETT-SI  
Bldg. 510-1-015  
CH-1211, Geneva 23  
Switzerland

[martin.vesely@cern.ch](mailto:martin.vesely@cern.ch)  
[www.cern.com](http://www.cern.com)

## Grey literature from invisibility to visibility: The quest for grey content in the domain of the invisible web

“I’ll make him Gandalf the White so he can really kick ass”<sup>1</sup>

Cees de Blaaij  
Zeeuwse Bibliotheek

### Introduction

In the second part of the trilogy Lord of the Rings: the Two Towers the wizard Gandalf the Grey becomes Gandalf the White. The change of colour is decisive for the further development of the story. Without becoming white and the help of Gandalf - having great powers - the notorious ring couldn’t have been destroyed at the end.

The motto I chose refers to the process of turning “grey” literature into “white” in terms of visibility and accessibility. At this moment we are on the threshold of a new phase in the development of the web. Second-generation web research and technology allow building of digital libraries that offer more and better bibliographic control. It looks very promising in solving problems of invisibility and accessibility of literature - grey included - in general if adaptation of second-generation tools and standards will succeed. With the realisation of the development of the web towards a ‘semantic’ web<sup>2</sup> it looks as if grey literature will become more easily to “discover”, and therefore is less grey or less obscure and turns white.

I want to address two questions and try to answer.

- The first question is in what ways is the availability of grey literature influenced by the evolution of the web in the last years?
- Secondly what about the accessibility of grey literature in the near future in the virtual world? I know these are a ‘big’ questions but I will try to outline present developments.

### Growth

The way we observe the evolution of the Internet nowadays has much in common with the perspective on the geography of the world at the end of the 15<sup>th</sup> century. At that time there was a lot of guessing about the geographical characteristics of our planet. Many parts of the world were still obscure and cartography was a bit casino like. The introduction and the usage of adequate navigation tools expanded the horizon for secure travelling. Explorations resulted in shifting boundaries of empirical geographical knowledge and colonization.

In contrast, we now talk about the earth as the global village. We have shrunk our planet virtually thanks to linked communication and information networks. The phrase “long distance phone call” is history. To some with a romantic vision a new virtual world was created parallel to the physical world and was hype named cyberspace. After the big shake out of dot.com companies at the end of the nineties we have come to a new phase of web space evolution. The World Wide Web - who could have ever predicted that - has become the most important function of the Internet.

It’s difficult to assess the size and growth of the web in general because different research methods have been used and so studies on this subject come with different results. In either case remarkable is the rapid ‘take off’ and fast initial growth. In 1996 it was estimated that the size of the Web as a whole was about 100,000 sites. Two years later there were nearly 1.5 million sites. By 2000, the public Web had expanded to 2.9 million sites, and two years later, in 2002, to over 3 million sites. In the five years it more than doubled in size.<sup>3</sup>

<sup>1</sup> Wizards: Mission impossible <<http://www.jrrtolkienpics.com/wizards.html>>

<sup>2</sup> The Semantic Web is a mesh of information linked up in such a way as to be easily processable by machines, on a global scale. You can think of it as being an efficient way of representing data on the World Wide Web, or as a globally linked database. For more information look to <<http://www.w3.org/DesignIssues/Semantic>>

<sup>3</sup> For numbers on size and growth look at <<http://wcp.oclc.org/stats/size.html>>

The growth of the World Wide Web in the last decade seems to have caused a kind of inflation of the value of grey literature due a kind of uncontrolled growth. Artus stipulated this at the last GL'99 when he mentioned that the so-called 'one-man institutes' and one-man-publishing houses would be responsible for that particular development.<sup>4</sup>

His estimation then was that we had to fear that grey literature had to be confused with an uncountable number of more or less worthless or at least "dubious papers". From that perspective it would be justified to say that grey literature would lose its reputation as a serious source for scientific information.

But fortunately the prophecy for such a scenario did not become true for different reasons. The original speedy growth can be explained from the fact that a lot of organisations rushed to be online. This did not continue. By the end of 1999 a lot of dot-com firms ended their Web operations or scaled back. The fear from Artus can be taken away concerning the threat of blurrification of grey literature. Analysts have noticed a decline in websites maintained by private individuals or "personal" Websites. This has to do with the fact that many free-of-charge Web hosting facilities have been stopped and many simply did not want to pay for the maintenance of their site<sup>5</sup>.

The most remarkable thing remains the rapid expansion of the web. But recently there has been a slow down in the growth of the public Web. An examination of year-to-year growth shows that in the period 1998- 2002 the growth of the public Web has been far less. Between 1998 and 1999 the public Web expanded more than 50 percent, between 2000 and 2001 the growth rate dropped to 6 percent and between 2001 and 2002 the public web became even smaller. On basis of the outcomes of the Web Characterizations Project Surveys one can draw the conclusion that the Web no longer grows. It has reached a kind of temporary equilibrium.

Another trend is that the OCLC's Web Characterization Project's data from June 2002 suggest that while the public web on the one hand is becoming smaller in the number of sites, the public websites on the other hand are getting larger. In 2001 the average number of pages per public website was 413. In 2002 this increased to 441 pages.

### **Role of grey literature**

As in regard to growth and the role of grey literature in the evolution of the web there has been some change in the ways grey literature has been electronically disseminated the last years:

1. On the one end there is the development the public web. By definition the public web is the whole of websites that offer free access to users.
2. Then there is a part of the web that is distinctive from the normal or surface web, the so-called "deep web" or invisible web. This particular part of the web is called invisible because it consists of material that is not indexed or included in the collections of the general-purpose search engines like Google and Yahoo. This does not say that information on the invisible web is inaccessible. Perhaps the term deep web is a better choice as there is nothing really invisible but needed content is not easily located. You have to know where to find it since you cannot use engines like Dogpile or Lycos

### **The role of the public web**

Pre-print servers disseminate a huge part of grey literature<sup>6</sup>. One of the big advantages is that electronic preprints allow access to information without the time lag inherent in traditional publishing. At this stage pre-prints in a number of cases are not peer-reviewed. A resulting weakness could be that the quality and validity of scientific results would be questionable.

---

<sup>4</sup> Helmut M. Artus, SMAISMRMILMEPOETALEUMBUNENUGTTAUIRAS. The Internet & the socio-structural change of informal scientific communication, in: GL'99, Conference Proceedings, p. 23

<sup>5</sup> G. Mariano, The incredible shrinking Internet < <http://news.zdnet.co.uk/business/0,39020645,2101890,00.htm>>

<sup>6</sup> Preprints are manuscripts that have not yet been published, but may have been reviewed and accepted; submitted but with no publication decision; or intended for publication and being circulated for comment.

On the other hand practice shows that the total number of print-print servers has continuously risen<sup>7</sup>. This shows that the scientific community has enough confidence in the way preprint servers fulfil their tasks and objectives. A number of these preprint servers operate as a part of the public web others belong to that part of the web that is called the invisible web.

According to the most recent results of OCLC's Web Characterization Project<sup>8</sup> the public web, as of June 2002 contained approximately 1.4 billion pages. In 1998 Shapiro and Varian<sup>9</sup> made an estimation that the static HTML text available on the Web amounted to 1.5 million books. They made a comparison with the number of books that were available in UC Berkeley Library and they concluded, "The Web isn't that all that impressive as an information source".

You can criticize this conclusion on the assumption that the web contains more than only static text. In the last years there has been a considerable development in the usage of different types multimedia files and other types of information objects. So if one assesses the size of the Web only on basis of static text then a lot of dynamic information is been ignored.

So it's better not to jump to any hasty conclusions about the relevancy of the web.

### The Invisible Web

Although many users think that the so-called invisible Web is not worth covering, being full of duplication, 'spam' and other ephemeral material, the fact remains that some of the best resources available on the Web exists in this form and to overlook it would be to detract from the full value of the Web.

Users should seriously try to discover the deep Web because of the large size of the resource, the quality of the websites in relation to a specific subject area. Specialized content could satisfy certain information needs from users. The major obstacle to using the invisible Web is the intrinsic difficulty in accessing it.

Many websites in the deep consist of normal WebPages. They can't be found by using a regular search engine for the simple fact that these websites are excluded because certain data types cannot be indexed.

Most search engines have difficulty indexing other formats than HTML pages. Compressed files, Flash animations etc cannot be handled. Lately search engines have improved their capabilities. For instance Google managed to index the format pdf en doc. For other search engines indexing this type of format could be very expensive. A pdf file can consist of thousands of pages. Besides that in the recent past the owners of the general-purpose search engines believed there was less demand for these types of files.

This argument seems to be outdated and is not a valid one. Many US and European government agencies en non-governmental organisations use the Portable Document Format as a de facto standard. In April 2003 the first draft of an international standard that defines the use of PDF for archiving and preserving documents on the Internet has been submitted to the International Organization for Standardization (ISO) for review.<sup>10</sup> Scholarly information is also produced in this format or otherwise in Postscript or related format.

Grey literature is not separate from other type of information. It comes from a great variety of sources like governmental and/or non-governmental scientific organisations and is indexed in special databases and are made available online through a web interface. These databases are often relational databases with sophisticated query tools to retrieve the required information. The only way to retrieve the right information is interacting with the database itself. It is this kind of content which constitutes the core of the Invisible Web.

---

<sup>7</sup> Nicholas G. Tomaiuolo, Joan G. Packer, Preprint Servers: Pushing the Envelope of Electronic Scholarly Publishing, in: Searcher, vol. 8, nr. 9, October 2000, p. 53

<sup>8</sup> Web Characterization OCLC Online Computer Library Center, Inc. Office of Research <<http://wcp.oclc.org/>>

<sup>9</sup> C. Shapiro and H. Varian, Information rules: a strategic guide to the network community

<sup>10</sup> PDF as a standard for managing and archiving documents and forms  
<<http://www.adobe.com/products/acrobat/pdfs/docsforms.pdf>>



Let's take a closer look to those types of invisible content and the reason why it can't be found:

Type of content	Reason for not finding it
Web page is not connected	No hyperlinks for crawlers
WebPages consist of file formats which can't be indexed like PDF, Compresses files etc	
WebPages consists of images, audio video	Technically indexable but in most cases ignored for business or policy reasons like potential copyright violation (mp3 files etc.)
Content in relational databases	Crawlers are not able to use query forms
Real time content / dynamical updated WebPages e.g. stock exchange inform	It is possible to mislead crawlers to "spamdex" the search engine (bogus pages). Use of "spider-traps"

Databases connected to the web often are difficult to index by search engines because they have a data structure of their own and a particular design. A search engine perhaps can locate the database but it can't get into the database itself and give you information about its contents.

So what about the future making the deep web more visible?

Search engines will become smarter. They will be better capable to make judgements about the relevancy of a page. Exclusions of websites will be less because computing power, data storage and bandwidth becomes less expensive so eliminating economic reasons for not digging deeper and indexing better. On the other hand manipulation of search engines will go on with different tools. So it will be an ongoing competition. I would not call this a dead end street because there is another important development going on.

The present state of the WWW is kind of equilibrium. HTML is still being used for weaving the Web. The problem of poor bibliographic control has not been solved because the used lingua franca<sup>11</sup> of the web - that is HTML - was not designed for semantic purposes as to give information objects a meaning. It was designed to function as standardized way to display documents and not meant as a navigation tool for content. The alternative to solve the issue for better bibliographic control in the case of grey literature was to index grey literature in special designed databases like SIGLE<sup>12</sup> and its Dutch counterpart GLIN<sup>13</sup>.

For both criteria: visibility and accessibility it is important that common standards are developed. Standards for managing electronic content are important to searching, functionality and the creation of stable electronic archives. For grey literature to become more and better accessible we need different publishing models build on the same standardized platform of scientific communication. I don't believe we should strive for a one-size-fits all publishing model for grey literature because the differences between several categories of grey literature are too big. What we do need is a more standardized platform for digital content in general.

<sup>11</sup> **Lingua Franca** is a trade language used by numerous language communities around the Mediterranean, to communicate with others whose language they did not speak.

<sup>12</sup> **SIGLE (System for Information on Grey Literature in Europe)** is a bibliographic database covering European non-conventional literature in the fields of pure and applied natural sciences and technology, economics, social sciences, and humanities

<sup>13</sup> 'Grijze literatuur in Nederland' (GLIN) is a Dutch bibliographic database containing descriptions of ca. 110.0000 publications, e.g. dissertations published in the Netherlands and publications from governmental en academic organisations. All of these are available at the Royal Library, The Hague, Netherlands

With this background picture it was welcome to upgrade HTML to a language of a higher level not only to the benefit of grey literature but for the content and structure of electronic documents in general

### The Semantic Web

Two technologies are available for developing what has been called the Semantic Web<sup>14</sup>: eXtensible Mark-up Language (XML) and the Resource Description Framework (RDF). RDF is a collaborative design effort; no one individual or organization invented RDF. Several W3C Member companies contribute intellectual resources to its development. The godfather of the Web Tim Berners Lee participated in the further development.

XML delivers an effective way to label information documents with metadata: data about data. Metadata provides standardized information about a document, including items as the name of the author, a summary and descriptive keywords. Metadata provide the capability to introduce a controlled vocabulary, which offers thesaurus like possibilities for searching. A controlled vocabulary can bring different terms and concepts together. Searches will become more effective because information objects like tables and figures will be able to be queried and to be used.

Meaning is expressed by RDF, which uses XML for encoding meaning in sets of triples, each triple being like a sentence with a subject, verb and object. The triples of RDF form webs of information about related things. In general, RDF provides the basis for generic tools for authoring, manipulating, and searching machine understandable data on the Web thereby promoting the transformation of the Web into a machine-processable repository of information. Important is to agree on common metadata standards for the basic structure of scientific documents. This will have an enormous impact on the accessibility and dissemination of scientific papers and grey literature in general. It opens possibilities to join big sources of grey literature like online journals, e-print archives en scientific digital libraries.

### Open Archives Initiative

Several initiatives have been established to promote common web standards. One of them is the Open Archives Initiative<sup>15</sup> that offers scientists the possibility to disseminate their work. Any group who likes to participate in OAI by creating an archive must agree to common standards become a part of the greater whole. By using a standard protocol collections provide descriptions about their contents. This is the basis for the creation of interoperable digital libraries. By standardizing the interface to the metadata, tools to work with this data have to be created only once. The benefit is that organizations that can't afford programmers can now have digital libraries, using these standard tools.

So its important that scientific communities work together even to make the adding of metadata economically viable. Adding metadata is expensive. By sharing and adding scientific papers in a community costs will be lower and retrieval of relevant information more rewarding.

---

### Author Information

**Cees de Blaaij**  
Zeeuwsebibliotheek  
Postbus 8004  
4330 EA Middelburg  
The Netherlands

C.deBlaaij@zeeuwsebibliotheek.nl  
www.zeeuwsebibliotheek.nl

---

<sup>14</sup> For information on semantic web <<http://www.w3.org/DesignIssues/Semantic>

<sup>15</sup> <<http://www.openarchives.org/>>

## 50 Years of Experience in Making Grey Literature Available: Matching the Expectations of the Particle Physics Community

Carmen O'Dell, David Dallman, Martin Vesely, and Jens Vigen.  
CERN

### Abstract:

The CERN Scientific Information Service has been active in the field of digital library research and in providing scientific information services to the high-energy physics community for almost five decades now. Most recently the research focus has been on interoperability issues in document storage and retrieval systems, metadata added-value services, digital library automation and networked information services. The achievements of this research and the implications for treating grey literature are presented, including practical implementation examples.

### Introduction

CERN, the European Organization for Nuclear Research, was founded in 1954 and is the world's largest particle physics laboratory. The Member States financing CERN are all European but the laboratory has researchers from all over the world. In fact, there are some 6500 CERN users, representing 500 universities and over 80 nationalities. CERN itself employs around 2500 staff to provide the necessary technical, computing and administrative infrastructure. The CERN Scientific Information Service was created in 1955 and one of its key tasks is to: "acquire and manage information resources in all fields of relevance to the Organization, and make these accessible in the most convenient way to the worldwide particle physics community" [1]. To do this effectively the library has continually adapted and evolved its methods for collecting, treating and disseminating information resources.

### 1. The Preprint Culture

In all fields of scientific research, there is a need to be the 'first' but research into the tiniest of particles ironically requires some of the biggest machines in the world, with price tags to match. When research is this expensive, there is simply no room to do the same science twice. Over 40 years ago this drove the particle physics community into a culture based on preprints, to accelerate scientific communication beyond that of the lengthy journal publication process. Universities and laboratories produced their own series of preprints describing their research and experimental results, and distributed them free of charge to hundreds of other institutes in the field.

In 1958 the CERN library started issuing a weekly list of the preprints received and by the early 1960s, semi-automatic means were being used to produce it. Eventually each document was catalogued and the bibliographic data typed into a database from which the list could be produced. In 1983 this database was made available to users for searching.

By 1991, around 10,000 paper preprints were being catalogued per year with publication references to the corresponding journal articles also being added to the database by hand. The system was absolutely at the limit of what could be achieved with the two staff members available. In addition, there were problems of space for storing all these paper documents.

Fortuitously, in August 1991, Paul Ginsparg set up the Los Alamos electronic preprint server, now known as arXiv. The server provided an enormous improvement in the speed and ease of preprint communication. By 1992, it was starting to become so popular among particle theorists

that it was not long before some institutes announced they were stopping the costly distribution of paper preprints.

CERN librarians realized that while they would benefit greatly from having less preprints to handle, the e-prints still needed to be integrated into the library collection in a meaningful way. A program was written to read the daily email alerts from arXiv, decipher the bibliographic data and create database records with links to the full text documents on the preprint server. Although the terms had not yet been invented, this was probably the first example of a program to harvest metadata [2] .

Manual input of paper preprints continued and most were scanned so that the library could offer an almost complete electronic library. The number needing to be treated in this way though decreased steadily as arXiv grew in popularity.

## **2. Digital Collection Building**

The library has used two different approaches to expand its digital collections beyond the data available at arXiv. One is library-driven and known today as metadata harvesting - other sources of relevant electronic documents are located and automatically uploaded to the bibliographic database. Once established, this takes place as long as the source continues to exist. The other is author driven with documents individually submitted to us via a Web form.

In both cases, if the fulltext of the document is available at the given source (not always the case, as we also harvest from databases/aggregators), our system automatically goes to retrieve it once the bibliographic record has been created. The digital document is then permanently stored on our server. When we consider that the source is at least as stable as ourselves, we may leave the full document untouched and provide an external link in our database. Clearly, this harvesting of digital documents and metadata only applies to material which is freely available.

### **2.1 Metadata Harvesting**

We wrote our first harvesting program in 1993 and since January 1994 it has been running on a daily basis. Originally we only harvested from arXiv, which still represents about 50% of all our documents. From 1998 onwards though, we steadily added new sources until the current number is over 100. Initially a new script was written for each source but when we had several such scripts, we saw that we were heading towards a maintainability crisis.

Since large parts of these programs were doing very similar tasks, we developed a program called the Uploader which can handle a wide variety of external data formats. The part specific to the individual source format has a simple low-level language in which the external format can be described. Writing a so-called configuration file for a particular external source can then be done by a librarian, without programming knowledge being needed. The types of external data that can be handled range from highly-structured bibliographic database records to essentially unstructured HTML pages [3].

Before records are added to the database they pass through a simple matching procedure using terms from the title and authors. If a match is found, any new bibliographic fields are added to the already-existing record. Otherwise a new record is created. This reduces doublet generation quite considerably.

The frequency of harvesting is very varied, and is largely determined by the service offered at the other end. If a profile can be established, one can harvest data according to a daily or monthly schedule. If there are just Web pages listing documents, one has to find a way to signal the occurrence of new documents.

Most of the sources from which we harvest are freely available, for others we have made agreements about what we are allowed to upload. For example, INSPEC (one of the major databases in the physics field), allows us to upload all records connected with work at CERN.

Use of standards would be an enormous help. We are heavily involved with the OAI (Open Archive Initiative), having been host to the OAI European workshops in both 2001 and 2002 and have been one of the European representatives on the OAI technical committee. Our own system is OAI compatible [4], but even if this was true of the external sources, the protocol has not yet been exploited to such an extent that we could download data in the complexity we need.

## 2.2 Submission Procedure

Originally designed for capturing in-house CERN documents, this procedure was started in 1996 and has grown in complexity ever since to include more and more of the documents produced by the laboratory. Unlike arXiv which has a deliberately simple submission interface, there are more than 30 web submission forms, each one tailored to a specific document series or type of information resource [5]. The submitter, who may be one of the authors or a secretary responsible for a document series, fills out the necessary bibliographic information in the relevant form and gives the path to the full text files.

We have also introduced the possibility for external sources to submit documents, following requests from some outside universities and research centres. These submissions are moderated to ensure that the submitted documents meet the scope and nature of our database.

If individual documents are located by us, the librarian may make a self-submission of the document. If it turns out that the source is regularly found to contain documents of interest, this may be the forerunner of a new harvesting configuration. It will depend on the quantity of interesting documents and the way the external source is structured. Although not planned as such, we also find that this submission procedure is taking over from the cataloguing module of our library system. It is especially useful for newcomers, who sometimes have little prior experience of cataloguing.

The submission tools are continually being updated and refined. We are able to handle documents that are confidential, access being allowed to only a specified group of people (such as a committee). Documents can be migrated from confidential to public status once they are released, by simply flipping a flag in the submission interface. We are also able to handle documents that have to pass through a refereeing procedure before they are made public. So far, we do not offer authoring tools: we expect a document to be complete by the time it enters the system.

## 2.3 Quality Control

On average around 800 notices are added to the database each week. This is far too many for each one to be manually checked but with author submission and automatic uploading, errors will of course occur. So a number of in-house checking programs have been developed to identify common errors in the data input. Using associated knowledge-databases some of these errors can even be corrected automatically. For the rest, the librarians work mainly on files containing only the fields concerned, an individual record will only be accessed in the very small number of cases where the corrections are more complicated. This system allows us to handle a large number of documents on a daily basis with a small staff and still maintain the quality of the data.

## 2.4 Beyond Preprints

The CERN Library has used various library management systems in the past. Since 1989 it has been ALEPH (Ex Libris, Israel), which we are still using for most internal library functions. In 1996 though we chose to develop our own user search interface (CDSweb) in order to offer the functionality we needed. CDSweb made it possible to cover many different digital collections with a single search interface [6]. Once it was launched, we could integrate other types of textual documents and even non-textual “documents” into the database. So we included the internal notes of some of the larger experiments, the notes used in the Large Hadron Collider (LHC) construction project, the documents of the committees responsible for approving and scheduling the experiments (proposals, minutes, status reports), press cuttings about CERN and many other textual documents. Non-textual collections included the CERN Photo collection, the database of particle physics research institutes, the CERN Historical Archive, videotapes, webcasts and even the objects contained in the CERN permanent exhibition (Microcosm). These items had not been handled by the Library before, either because of the lack of resources for doing so in the paper era, or because they were considered outside the scope of the Library when the database was simply the library catalogue

As CDSWeb expanded into these other areas, the submission procedures were expanded in parallel, so that each type of document could be submitted directly to the database by the authorized people.

## 3. Linking to the Digital Documents and Other Information

### 3.1 Metadata Links

The most basic link is from the metadata describing a document to the document itself. For the electronic preprints, the URL has been stored as part of the metadata at harvest time. The corresponding article in an electronic journal is treated differently though since storing the URLs of every article in our database is not a scalable solution. Instead we developed our own link manager, GoDirect, which takes advantage of the fact that in many cases the URLs for articles can be derived from the journal title, volume and page numbers. From the journal reference present in our metadata, GoDirect can create a link on the fly to the actual journal article [7].

We were able to convince other publishers to follow this example of a “friendly URL” but some publishers were only willing to offer us files of correspondences between volume/page and the DOI. This solution is unsatisfactory though as it means data has to be stored for each electronic journal containing the URL format and the volumes that are electronically accessible. This has since been solved by CrossRef, an initiative from a consortium of publishers which provides a guest interface that enables DOIs to be resolved from journal/ volume/page information.

Another useful link is to pass from our metadata to the metadata of the same document in another database. Whenever we harvest metadata from another database we store the system number of that record (if it has one) in our own database. Linking to a record in another database may give access to additional digital full text versions. For example, the particle physics laboratory KEK in Japan has made an enormous effort to scan paper preprints from the pre-electronic era. Linking to another database can also be useful if our own files become corrupted in some way.

Another possibility is to pass from one set of documents to a group of related items, by automatically generating a database search from the metadata. For example, we can go from a conference document to all other documents relating to the same conference, ask for all the publications of one of the authors or search for more documents having the same subject

descriptors. We also have links from our metadata to non-bibliographic information resources in our field, enabling users to access explanatory or numerical information.

### 3.2 Citations Linking

For all electronic documents on our server (now about 250,000) we have extracted the block of references at the end of each article and have indexed the complete text of each reference [8]. For practical reasons, these have been stored in a separate citations database linked to the main bibliographic database on a record-by-record basis via the e-print number. In this sense, the citations have also become additional metadata.

This database permits the making of citation searches using any text which is in the citation. Additionally, the user can choose to display the list of references for a document. As this page is assembled, links to all e-journal references to which we have access are established using our link manager, as well as any links to e-print archive documents.

There are about 3 million linkable references inside our electronic documents. The handling of the citations in this way is carried out automatically for all new documents. A knowledge database with some 3000 entries is used in the standardization of the presentation of the citations. The project was started in 1998 and further improvements are underway to increase the proportion of references that can be linked to the full text.

This feature means you do not need to pass back through our metadata in order to arrive at the digital document. Apart from the obvious fact that it takes one click less to get to the full text, another huge advantage is that a document can be linked to even when it is not in our database at all. Thus when more than 100 years of *Physical Review* were put online (back to 1893), all the links to these articles immediately became available in the citations of our documents, even though we do not have the metadata for the older articles in our database. However, one big disadvantage arises from the fact that (for obvious reasons!) the journal article never contains any reference to an e-print. Thus, unless the author has also referred to the e-print as well as the journal reference, the e-print version will not be accessible from the citation list. If the user does not have access to the e-journal in question, they will then not have access to the digital document. There is also the reverse case when the original citation is only to an e-print because it had not been published in a journal at the time it was cited. The citation stays like this forever with no direct link to the later article.

## 4. Future Developments

### 4.1 Full Text Indexing

The full text of all our electronic documents has been indexed using the Internet search engine UltraSeek. This allows any text in any of these documents to be retrieved. This full text indexing is also part of our daily processing of new documents. At present it is a standalone feature so cannot be combined with searches in the bibliographic data. However this should be possible in the next version of CDSWeb which will be released shortly. In this case one could say that the whole text of the document has become metadata!

We are also planning to create permanent Web pages for all of our documents, so they will be picked up by the various Internet indexing programs, and hence by users completely outside of the library website. Conversely, we want to make it possible to launch an Internet search from our metadata. Very often this is a successful way of locating a digital document when all else has failed, for example when the document exists on the author's own Web site.

## 4.2 Automatic Keywording

We are using statistical and linguistic analysis of the complete full text in order to automatically attribute keywords and keyword phrases to the documents [9]. We have made a correlation between the full texts of a training sample of about 2000 documents and the thesaurus terms attributed to those documents by the documentation group at our sister laboratory DESY in Hamburg. From CDSWeb, it is now possible to automatically generate a list of DESY Thesaurus terms for any document. This project is currently in its first phase and more development is necessary.

## 4.3 Lexi

We are developing an encyclopaedic database of all terms used in the particle physics field (20,000 at present). When finished, we plan to link the mention of any of these terms (except perhaps the very common ones) in the full text of any document to a description of the term, with further links to the original documents that defined it.

## 4.4 GRACE

CERN is a partner in GRACE (Grid Retrieval And Categorization Engine). This project is developing a search engine that will allow users to search through heterogeneous resources stored in geographically distributed digital collections. GRACE will be run on the European Data Grid and will not have a centralized index as current search engines do [10].

## Conclusions

Currently about 95% of the particle physics literature is available to us electronically. By treating paper and digital documents in essentially the same way from the very beginning, the transition from a paper-based to largely digital library was able to take place seamlessly at its natural pace, as more and more documents became available in digital form. Since 1994 we have moved from a collection of paper documents described by a searchable computer-based library catalogue, to a largely digital library. The scope has expanded enormously to take in types of document that were not present in the traditional library.

The links to the e-version of the preprint and the corresponding electronic journal article stand side-by-side. This is especially useful for the non-CERN users who won't have access to the e-journals via our site. In the year 2001, queries were received from about 150,000 different host computers around the world.

We have developed a lot of functionality in the way we can access our documents since the start of the electronic era, despite the fact that the Library staff has been reduced in this period. As is often the case, automatic techniques do not really save time, they just enable one to do more in the same time. The database contains around 500,000 documents and is now expanding at about 50,000 documents per year, five times more than we could manage in the paper era.

The electronic document era has changed *how* we do things quite a lot, but so far it is hardly changed *what* we do at all. Despite the success of preprint servers, authors still need to publish in journals which while they may be electronically accessible, are still essentially clones of the paper versions. Scientific communication still follows the path of producing completely contained documents containing an introduction and a description of the context in which the work is to be seen, only a part of the document deals with what is really new. But the possibilities for accessing and linking documents via widely different types of information continue to grow. When scientists decide to move beyond the limitations of the paper-



publishing paradigm and exploit the possibilities offered by the digital age to the full, then the real scientific communication revolution will begin.

## References

1. CERN Scientific Information Service Mission Statement  
[http://library.cern.ch/library\\_general/group\\_mandate.html](http://library.cern.ch/library_general/group_mandate.html)
2. Dallman, D. 'Electronic Journals and Electronic Publishing at CERN: A Case Study'. *Intern. Spring School on the Digital Library and E-publishing for Science and Technology, CERN, Geneva, 2002.*
3. Pignard, N et al. 'Automated Treatment of Electronic Resources in the Scientific Information Service at CERN'. *HEP Libraries Webzine, Issue 3/Mar 2001*  
<http://library.cern.ch/HEPLW/3/papers/3/>
4. Vesely, M. et al. 'Creating Open Digital Library using XML: Implementation of OAi-PMH'. *International Conference on Electronic Publishing, ELPub2002, Karlovy Vary, Czech Republic, 2002.*
5. CERN submission interface visible at: <http://documents.cern.ch/EDS/current/access/index.php>
6. CDSweb accessible at: <http://cdsweb.cern.ch/>
7. Vigen, J et al. 'Link Managers for Grey Literature'. *4<sup>th</sup> International Conference on Grey Literature, Washington, DC, USA, 4 - 5 Oct 1999.*
8. Claivaz, JB et al. 'From Fulltext Documents to Structured Citations: CERN's Automated Solution'. *HEP Libraries Webzine, Issue 5/Nov 2001.* <http://doc.cern.ch/heplw/5/papers/2/>
9. Le Meur, JY et al. 'Automatic Keywording of High Energy Physics'. *4<sup>th</sup> International Conference on Grey Literature, Washington, DC, USA, 4 - 5 Oct 1999, pp. 230-237.*
10. Haya, G et al. 'Developing a Grid-Based Search and Categorization Tool'. *HEP Libraries Webzine, Issue 8/Oct 2003.* <http://library.cern.ch/HEPLW/8/papers/1/>.

---

## Author Information

### Carmen O'Dell

CERN Scientific Information Service  
The European Centre for Higher Energy Physics  
ETT-DH CH-1211, Geneva 23  
Switzerland

Carmen.O'Dell@cern.ch  
Martin.Vesely@cern.ch  
www.cern.com

## **GL systems and services in the specific fields of vocational training and labour policies in Italy: the ISFOL case**

Isabella Pitoni and Diana Macrì  
ISFOL

The recognition of the central role that issuing organisations have in the production, distribution and management of their grey literature is the theoretical context for this paper.

The paper focuses on a particular kind of GL system settled by an Italian public research institute primarily involved in vocational training and labour policies, in the framework of the EU strategies for the European Social Fund. Within this context, it aims to present the GL systems and services that ISFOL makes available through its Specialised Documentation Centre (SDC) in tight co-operation and collaboration with the European Union and the Italian Ministry of labour and social policies.

We focus on how ISFOL GL policy is changing, on one hand thanks to the development of communication and information technology and on the other hand following European Union policies for public access to documents and for the development of the e-Europe and the e-government.

Grey literature produced by ISFOL and the SDC, often in co-operation with the Italian Ministry of labour and social policies, is analysed. At the same time a description of the system of related websites where all ISFOL grey literature is made available online is carried out, to show how new technologies and policies influence the Institute's strategies for making GL available for researchers and operators.

In conclusion, suggestions and contributions are made to create a roadmap of an integrated GL system involving organisations operating in the field of vocational training and labour policies.

### **Introduction**

The definition of grey literature known as Luxembourg Convention and formulated during the 3rd International conference on grey literature in 1997 can be a good way to start this paper. It says: "GL which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publisher"<sup>1</sup>. Compared to the former "official definition" of grey literature given during the Seminar of York in 1978<sup>2</sup>, although still stressing how GL's main characteristic is being outside the control of commercial publishers, it implies two important differences: on one hand the recognition that it is always within an organisation that grey literature is produced, and on the other hand the inclusion of electronic documents as possible form of GL.

The first aspect of the Luxembourg definition is not completely new. Although the importance of the issuing organisation has been a characteristic of GL since its appearance, today, thanks to the development of information and communication technologies, these are also involved in its control, management and distribution.

The importance of issuing organisations for GL production has been also at the centre of the latest discussions on grey literature in Italy. Of special interest was the presentation that

---

<sup>1</sup> Perspectives on the design and transfer of scientific and technical information. 3. International conference on grey literature, Luxembourg, 13-14 November 1997, Amsterdam : Transatlantic, 1997, p. III.

<sup>2</sup> "Il concetto di "letteratura grigia" equivale in linea di massima a quello di "letteratura non convenzionale". La caratteristica principale di questo tipo di letteratura, la cui pubblicazione in taluni casi non è prevista fin dall'inizio, è quella di non essere diffusa tramite i normali canali di pubblicazione commerciale, e quindi di essere difficilmente accessibile". La definizione concordata dagli esperti partecipanti al seminario di York è poi integrata di un elenco di esempi di "letteratura grigia" (In Alberani, Vilma, La letteratura grigia, Roma : La Nuova Italia scientifica, 1992, p. 28-29).

Alessandro Sardelli has given during the 3rd National conference on grey literature that took place in Rome in November 1999<sup>3</sup>. He proposed to add to the traditional expression “grey literature” the new one of “organisation literature”<sup>4</sup> as is always a strict link with its producer that, according to Sardelli, characterize GL in its different forms. He has also singled out four different areas in which grey literature can be applied: 1. Technical and scientific research field; 2. Commercial and productive area (often strictly related to the previous one); 3. The world of non-profit organisations; 4. And the public sector.

All these types of organisations, although very different in their institutional missions and targets, produce information and documentation that is in many case printed and distributed outside the traditional commercial publication channels.

For our purposes we are particularly interested in Sardelli’s observations on public organisations. He recognizes how it is within the public sector that GL finds its ideal means of expression intended as a “vehicle of direct communication, of participative democracy, of organisational transparency and of recovery of of historical memory”<sup>5</sup>. If grey literature was born initially in a technical-scientific research context, it later developed also in other disciplinary areas and found there new means of expressions and new forms.

The example of GL produced by the Institute for the Development of Vocational Training for Workers (ISFOL) is interesting because it puts together the public role of the institute with its operative research activities in the field of vocational training and labour policies. ISFOL, because of its institutional role, is interested in producing documents able to quickly reach a wide range of users, from training-system operators, to Public administrations, Social partners, researchers and students. But at the same time, because its institutional role, it wants to stay outside the traditional commercial publication channels.

The other aspect of the Luxembourg convention definition that we are considering, related to documents in electronic format, must be considered in the wider context of the impact that technological innovation had and is still having on the world of communication and information and consequently also on grey literature.

Internet in a very short period of time has developed as a pre-eminent means of distribution of information and documentation (a part of if it is grey or conventional literature). But especially for GL it represents an ideal means of distribution, as it contributes to solve the problem traditionally related to GL: its little visibility and difficulty of its access.

At the same time, if Internet solves some important problems related to visibility and accessibility and makes distribution of documentation faster and more efficient, it also creates some new questions such as validation, quality, durability and copyright.

Internet and the new digital technologies, besides contributing to improve the speed and efficiency of distribution of traditional GL, have spawned the development of new forms of documentation that have been called “grey literature of new generation” or “grey information”. If on the Internet we find the electronic version of technical and research reports, bibliographies, translations, conference papers, dissertations, etc., today there are also “new forms of GL represented by discussion groups, FAQ collections, e-journals and e-zines etc., new forms of grey documentation in its most traditional and well established meaning of “material not spread through the normal commercial publication channels”<sup>6</sup>.

These new forms of grey documentation carry on the same problems of all information available in Internet. These problems include selection, valuation, individuation and control. Also because this “grey literature of new generation” is not included in traditional databases and bibliographies of GL it is often difficult to know it exists.

---

<sup>3</sup> Sardelli, Alessandro, “Il fascino discreto della letteratura grigia: ridefinizione e ambiti di applicazione” in *La letteratura grigia: politica e pratica*. 3° Convegno nazionale, atti a cura di V. Alberani e P. De Castro, Roma, ISS, 2000, p. 15-20.

<sup>5</sup> Op. cit. p. 18

<sup>6</sup> Alberani, Vilam - De Castro, Paola, *Grey literature from the York seminar (UK) of 1978 to the year 2000*, ISPEL 35(2001), p. 245

On the other hand, compared to the enormous quantity of documentation made available online every day without any control, this kind of documentation offers guarantees of quality and the opportunity of being singled out exactly because it is produced within an organisation.

The official website of an organisation is today the main place where to find information on the its activities, projects and the documentation it produce. It is also often the best place to start a research on a specific topic for which the organisation is responsible.

Today most of organisations use their website as a preferential means for providing information on its publications and GL and often (when not restricted by copyright impediments) they make them available online (availability can vary from basic bibliographic information, to indexes and abstracts, up to the full text of the document).

Internet and new technologies offer a lot of new opportunities and today many organisations producing GL are also engaged in its management and distribution. Through special libraries, documentation centres or publishing areas they often take care of retaining, making inventories, cataloguing, indexing and distributing their documentation. Often they implement online catalogues and databases specialised in documentation produced at national and international level on topics for which they are responsible (extremely useful tools for specialised research). It is then clear how information and communication technologies offer new opportunities to organisations traditionally producing grey literature. This is even more true of research public institutions as one of their institutional tasks is providing free access to the results of their research.

ISFOL finds the tools provided by information technologies as essential instruments for accomplishing its institutional tasks, especially those related to communication and distribution of information and documentation on its activities and research. The role of documentation and its distribution acquire a growing importance and specific projects such as the Specialised documentation centre (SDC) bears witness to the fact.

### **The development of European Union policies: from grey literature to the public access to information**

The ISFOL policy of access to information and documentation can be better understood in the wider context of European policies on the same subject. This is why it is useful to provide an outline that goes form the European activities on the specific topic of GL, starting from the organisation of the York Seminar in 1978, through to the recent European strategies for information and communication.

As is well known, EU interest (then European community) in GL dates back to 1978 when, together with the British library, it promoted the Seminar of York on non-conventional literature. Since 1978 what generated EC interest in GL were themes related to bibliographic control, in order to guarantee availability and accessibility of GL within EU countries (the same themes were at the centre of attention, at the time, also for conventional material)<sup>7</sup>.

Two years later the European Union engagement produced the implementation of the SIGLE (System for Information on Grey Literature in Europe), a bibliographic database initially covering GL produced in the technical-scientific field but later enlarged to other disciplines such as economy, social science and humanities. SIGLE was created mainly to provide access to documents and incidentally to improve bibliographic coverage.

At the end of EU financing in 1985 a new non-profit organisation was created by member states to manage the SIGLE, the EAGLE (European Association for Grey Literature Exploitation).

In recent years, in the framework of activities for the development of the "Information society", and specifically within the e-Europe action plans, the EU recognised the primary role of what has been called "public sector information".

European Union dealt with the subject first in the "Green paper on public sector information in the information society" and more recently with the communication from the Commission of the European communities "E-europe 2002: creating an EU framework for the exploitation of public

---

<sup>7</sup> IFLA and Unesco were discussing and promoting the Universal bibliographic control (UBC) and the Universal availability of publications (UAP) projects.

sector information” that has developed into a “ Proposal for a Directive of the European parliament and of the council on the re-use and commercial exploitation of public sector documents (presented by the Commission)”.

To explain the meaning of the expression “public sector information” we find help in these European Union documents. They clarify that for their purposes information is “any content whatever its medium (written on paper or stored in electronic form or as a sound, visual or audiovisual recording)”<sup>8</sup> and that when they speak of public sector, “public bodies at different levels of government, central, regional and local, shall be taken into consideration. In all cases, state owned companies operating under market conditions and subject to private and commercial laws are not meant to be covered by either of these definitions”<sup>9</sup>.

A list of examples of this kind of documentation is also given: “Financial and business information is collected by a number of Ministries and public sector organisations. Company registers, required by law in many Member States, are maintained by the public sector. Legal information (in particular concerning legislation and jurisprudence) and administrative information is public sector information par excellence. Patent offices are usually public sector bodies. Scientific, technical, cultural and medical information is extensively collected by public research institutions and public archives. Geographical information relevant to transport and tourism (maps, road traffic situation) is also available in public sector agencies. Tourist information is gathered and published by public sector bodies at different levels of government”.<sup>10</sup>

These definitions clarify how public sector information is all documentation produced and distributed by Public administrations or also “grey literature of public sector”.

Even if EU documents focus on the economic and internal aspects of public sector information they recognise its value in the development of democratic and civil life within the EU.

They recognise how public sector Information / documentation acquire a growing value in the context of the “Information society” and of the development of new technologies. As a matter of fact, information and communication technologies facilitate diffusion and exploitation both for commercial and non-commercial purposes. In this context information can contribute to the promotion of economic and social development of the EU and facilitate circulation of goods, people and services within the EU. To make it possible, EU documents stress the need of a harmonisation of the different legislation in the member states.

Transparency, accessibility and openness are the basic principles that can be highlighted in the EU documents on public sector information. These principle should inspire the action of the public administrations in the member states.

### **ISFOL and the SDC and its policies for production and management of GL**

ISFOL, as a public research institution active in the framework of EU activities for European social fund is careful to adopt the fundamental principles that characterise European Union policies in the field of information and documentation management.

ISFOL, created in 1973, is a state-owned scientific research institution engaged in promoting and developing vocational training and labour policies in Italy. It carries out its activities in collaboration, not only with the Ministry of labour and social policies, but also with Regional authorities, Social partners, State departments, the European Union and several international organisations.

Between its institutional tasks we find the promotion of study, research, documentation and information activities and the exploitation, diffusion and transfer of the results of its activities.

---

<sup>8</sup> Communication from the Commission to the Council, the European parliament, the Economic and social committee and the Committee of the regions: e-Europe 2002: creating a EU framework for the exploitation of public sector information, COM (2001) 607 final, 23.10.2001

<sup>9</sup> Green paper on public sector information in the information society, COM (98)585 final, adopted on 20 January 1999 COM (2002) 207 definitivo - 2002/123 (COD)

<sup>10</sup> Communication from the Commission to the Council, the European parliament, the Economic and social committee and the Committee of the regions: e-Europe 2002: creating a EU framework for the exploitation of public sector information, COM (2001) 607 final, 23.10.2001

Over the years, the institutional objectives of ISFOL have developed in harmony with the social, political and economic changes taking place in the country.

Since 1995, the Ministry of labour has requested ISFOL to set up the Technical Assistance Structure (TAS) for the European Social Fund under the European Social Fund 1994-1999 Programming. The Structure's task is to provide support and assistance to the Ministry of Labour and Regional Authorities in managing and implementing actions co-funded by the ESF and envisaged in the Multi-regional and Regional Operational Programmes for Objectives 1, 3 and 4. ISFOL today actively operates within the framework of the European Union strategies for European social fund.

The Specialised Documentation Centre (SDC) is one of ISFOL's projects set up in 1997. It is a Centre for specialised multimedia documentation on themes for which ISFOL is institutionally responsible and is intended to provide theoretical as well as technical and operational documentation answering the needs and interests of national training-system operators. Specifically, it produces and distributes innovative and integrated documentation products and services to support social and economic research and the design and management of training activities.

The SDC also operates within ISFOL's Structure for Technical assistance to the Ministry of Labour and Social Policies UCOFPL (Central Office for Vocational Training and Guidance) and it receives funding for these activities from the European Social Fund.

The SDC carries out several activities of documentation and management and promotion of documentation produced by the centre and by all of ISFOL. It collects, catalogues, indexes and makes documents (often GL) on vocational training and labour policies available to training operators, researchers and citizens.

Most of the documentation produced by the Institute, especially the one produced in the framework of EU strategies for the ESF, is not published through commercial publishers and is available for free. Often it is also available online in electronic version on the ISFOL website ([www.isfol.it](http://www.isfol.it)) or in the Eurolavoro web pages ([www.welfare.gov.it/Eurolavoro](http://www.welfare.gov.it/Eurolavoro)).

Many are the specific activities of the SDC related to the production, management, control and diffusion of GL of public sector produced by ISFOL:

- Creation and implementation of the Catalogue of ISFOL documentation (which includes grey and published materials). It is an electronic database of all documentation produced by different ISFOL sectors since 1974.
- Implementation and updating of the LOGOS database, which includes bibliographic records of documents on vocational training labour policies and EU activities for the ESF produced since 1998 and collected by the CDS.
- Co-ordination of editing and publishing of monographs, serial materials, and multimedia products both at institutional level and in the framework of its activities for the European social fund.
- Participation to the SIGLE system (System for Information on Grey Literature in Europe) collecting the bibliographic records of GL produced in the Institute and providing details to the CNR central library national centre for Italy.
- Dissemination of information on ISFOL and ESF activities and distribution of its products through the Internet.
- Management of the CDS front desk open to the public, and of a data service for enquiries and thematic reports on request of Institutional bodies.
- Management of a specific documentation system aimed at promoting information on European policies, with particular reference to the ESF.

In the context of these activities, the SDC is engaged in guaranteeing bibliographic control and availability of GL produced by ISFOL with different strategies. We can mention the attribution of the International Serial Standard Number (ISSN) to all periodicals and serial material produced by the Institute, a fundamental tool to allow its identification at international level.

In addition, the recent initiative for obtaining the status of non-profit editor that will guarantee to ISFOL documentary production a better visibility and bibliographic control and at the same time the opportunity to still make it available on the Internet without problems related to copyright restrictions.

Participation to the SIGLE system since 1998 (just after the CDS project started) is also indicative of the recognition of the scientific and informative value of grey documents and of the importance of its standardisation, control and availability.

The production of periodicals of documentation, in particular the bibliographic bulletins EUROPA.DOC e EUROPA.DOC special, also represent a contribution to promote bibliographic control and diffusion of documents produced (mainly grey literature) on specific topics. In particular EUROPA.DOC focuses on grey documentation produced by the European Union institutions and agencies and by research organisations at a European level operating in the field of vocational training and labour policies.

### Traditional and “innovative” grey literature in the ISFOL and CDS web pages

From the analysis of the ISFOL website and of the CDS web pages on the Ministry of Labour and Social Policies website, it is possible to highlight strategies for management and distribution of traditional grey literature and at the same time the emergence of “new forms of GL”.

To find GL produced by the Institute in the ISFOL website (<[www.isfol.it](http://www.isfol.it)>) we need to access the “Publications” page. In it we find a list of published materials and grey literature produced by different areas and projects of the Institute in the context of their institutional activities. Documents included in the list are all printed materials but they are often available also in electronic format.

Grey documents (not produced by commercial publishers) described in this page are:

- **Osservatorio ISFOL** Indexes of all issues are available online.
- **Monografie sul mercato del lavoro** Monographic series on the labour market. Full text of all documents are available online. Also abstracts in Italian and English are available.
- **Flailab web news** Newsletter on the Flailab project. All issues are available online in full text.
- **INO (Isfol notizie)** Monthly newsletter on ISFOL institutional activities (produced by the SDC). All issues are available online in full text.
- **Europadoc e Europadoc speciale** Bibliographic bulletins on European community documentation and on Social dialogue in the South of Italy (produced by the SDC). All issues are available online in full text.

But this is not all the GL produced by the institute and visiting other pages of the website it is possible to find other products.

In the web page of the Ufficio stampa (Press office) we find, for example, information of the existence of the internal **ISFOL newsletter**, a bimonthly newsletter edited by the Press office intended for internal use. It aims to inform Institute researchers and personnel about the activities of different areas and projects. It is not available online because of its character as an internal work tool.

An example of a rich source of GL information are the pages of the area “Sistemi informativi” (Information Systems) that has a page dedicated to “Publications and documents”. In it there is a variety of GL products not listed in the main publication page. Here we find, between other things **Reports on vocational training** offered in Italy and **Statistics of vocational training** all of which are available online in full text.

From this analysis we have seen how a consistent part of GL produced in printed version is available also in electronic format. ISFOL in fact, when is not bound by copyright, makes its documentation products available online. On the other hand, published documents are not available on the website (we only find indexes and abstracts).

Whilst recognising the importance of Internet for dissemination of information and documentation, ISFOL is aware of the important role as an information source that traditional printed documents still carry for many people. For this reason electronic documents have not replaced printed materials. Different strategies are adopted at the same time to carry on in the best way the missions of the Institute related to dissemination of information and documentation.

Next to the electronic version of “traditional grey materials” on the ISFOL website we also find other documentary products that can be defined as “grey literature of new generation” and that carry on an important and innovative informative role.

Examples of these types of GL on the ISFOL website accessible from the homepage are:

- **Institutional information:** pages providing information on the Institute.
- **Bacheca (Showcase):** periodically updated relevant information. Contains also information on the latest documentation produced by the institute, eventually with a link to the full text (for example the latest issue of INO Isfol or the latest monographs published)

From the section for “Research and projects” it is possible to access the pages of the different areas and projects that are a rich source of “grey information”.

Besides detailed information on activities carried out, ongoing projects, documentation produced and contact details, we often find in these pages links to legal information of interest for the area or project (for example the “Legislation” page in the Continuous training area), access to databases (examples: Legislative archive ARLEX in the web pages or the Orfeo Informative system), links to websites of national and international organisations co-operating with ISFOL, etc.

Other examples of grey documentation are available in the pages of the Flai Lab project where we find a **Map of business opportunity** and a guide for new entrepreneurs **Voglia di fare** both online in full text.

We will examine now [www.welfare.gov.it/EuropaLavoro](http://www.welfare.gov.it/EuropaLavoro) the website of the Central Office for Vocational Training and Guidance (UCOFPL) of the Ministry of labour and social policies where one can also find the SDC webpages.

In the section called “Communication products and services” we find a list of links. The first “Publishing products” contains all documentation produced by the CDS together with other documents produced in the framework of the European social fund. All documentation produced within the strategies for the ESF is entirely grey. This is because the rules of the ESF impose the wider dissemination and free access to products financed by the fund.

Documentation available is:

**Eight series:** FSE obiettivo 1; FSE obiettivo 3; Vigilanza e controllo; Alla scoperta delle professioni; Vivere l’Europa; Il lavoro stagionale; Guide stage; I libri del fondo sociale europeo.

**Three Newsletter:** FSE News; FSE mail; Europolavoro vi informa.

**Three periodicals:** Europa.doc; Fop; Risorse news.

**Five Brochures:** Forma lavoro; Fare impresa; Formazione; Inclusione sociale; Ricerca del lavoro.

**Other publications:** Manuali di comunicazione; Analisi dei fabbisogni formativi in agricoltura; Agricoltura ed ecoambiente; Modelli orientamenti appennino; I titoli che valgono un lavoro; Settori in cerca di lavoro.

The different categories of grey documents covered are monographs, pamphlets, brochures, periodicals, newsletters, electronic sheets, etc. All are available online in electronic format. Between the documentation produced by the CDS, apart from **Europa.doc** (available also on the Isfol website), there is the monograph “I termini della formazione: il controllo terminologico come strumento per la ricerca” published in the series “I libri del Fondo sociale europeo”. It contains the **CDS thesaurus**, a controlled indexing language formed by a list of terms on vocational training, guidance and labour which is also used as a work tool for indexing in the CDS databases.

Also in the Europolavoro website a variety of “grey information” is available. The **FSE mail** service is an interesting example because is listed as a traditional document. FSE mail is a service providing updated information on ESF via email that fully exploits the continuous updating potential of Internet.

Another link, still in the section “Communication products and services” called “Multimedia products”, gives access to the online version of some CD-ROMs produced in the context of the ESF.

In the pages of the SDC, still in the same section, we find detailed information on the project and on its activities and services (including documentation products). Here it is possible to access the two databases implemented by the SDC: the catalogue of ISFOL documentation and LOGOS databases.

In the Europolavoro website there are also many other information resources that can be included in the category of “public documentation”. Detailed information on activities related to the ESF in Italy, on training guidance and on the European community initiatives and



programmes. For example we can find catalogues of EU and national legislation related to the ESF and a glossary of terms related to the ESF programming.

### **Open problems, conclusions and a concrete proposal**

From the map that we have outlined, it can be seen how grey literature produced by ISFOL and its SDC is very rich and varied and constitutes an important informative resource. At the same time the SDC, exploiting opportunities offered by the Internet and IT technologies, provides a wide variety of GL services that are a fundamental point of reference for training operators, researchers and citizens interested in topics related to vocational training and labour policies within the EU strategies for the ESF.

However, at the same time we have seen that there are many problems still open both at a national and international level. If tools set up to guarantee bibliographic control and availability of GL, at local and international level, have developed thanks to new technologies, at the same time new problems have arisen especially in relation to new forms of grey documents that are mainly online (problems that are not specific for GL but that involve all kind of online information and documentation).

The enormous quantity of new information and documentation available everyday on the Internet makes its control and access even more difficult than in the past. The answer that is increasingly often given to these problems is the creation of smaller and more specialised archives and databases (compatible with international standards).

In the field of vocational training and labour policies, with special reference to grey literature, the creation of a specific database collecting traditional and “new” grey literature produced in Italy by public and private organisations would be useful. The SIGLE system should remain anyway a point of reference, both for the method of collecting data (based on inter-institutional co-operation) and for the record format that should be compatible with it.

Organisations producing grey literature in a specific field should be identified and involved in the project and ISFOL, because of its institutional role at a national level, could be the co-ordinator of this project through the SDC.

---

### **Author Information**

#### **Isabella Pitoni**

ISFOL

Visite di studio per specialisti della formazione

VIA MORGAGNI 33

00161 Rome

Italy

[i.pitoni@isfol.it](mailto:i.pitoni@isfol.it)

[d.macri@isfol.it](mailto:d.macri@isfol.it)

[www.isfol.it](http://www.isfol.it)

## Emergent Recognition: An Alternative Perspective on the Grey Literature of an American Public High School, 1985-2001

Hal von Hofe  
Horizons

### Abstract

The grey literature produced by a social body records and reveals interactions and communications which would otherwise remain only vaguely remembered or not at all: it offers alternative (plural) forms of memory for analysis. The cybernetic apparatus linked to the 'telephone lines' has greatly facilitated and amplified the production of GL.

The effects of the introduction of the cybernetic apparatus (computers, internet) on an alternative school program ("Horizons") within a specific social body, Branford High School, an American public school serving a small and relatively affluent shoreline community of about 30,000 in the Connecticut, will be analyzed here in terms of its grey literature production across a 16 year period (from 1985-2001).

Branford High School itself services approximately 1,000 students a year, with a teaching staff of about 80, close to a dozen administrators (with secretaries), a maintenance staff of around 6. The "Horizons" program, operating within Branford High School, services some 60 of these thousand students a year, those who have been identified as "at risk", or as "disaffected learners" (two terms of many from its grey literature). It forms a sort of school within a school, and presently consists of 8 teachers, an administrator and a social worker. -- I worked as a teacher of English (language and literature) in this program during that time, and offer my observations from that perspective.

Following the basic definition of grey literature, as given in the program guide of this conference: "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing" - I offer, as theoretical framework, these four levels of discourse within the literature of Horizons and Branford High School:

- 1) of the students (product/clientele/industry)
- 2) of the teachers (knowledge/expertise/academia)
- 3) of the administrators (business)
- 4) of the law (juridico-legislative, government).

I link these further with the four discourse model, and attendant schemas and algorithms, proposed by the psychoanalyst Jacques Lacan, as elaborated by Dr Theaux - a model to which we shall return, following a brief introduction to and history of the "Horizons Program" in Branford.

The "Horizons Program" began originally under the name of "Basic Studies", in 1973, with 4 teachers and one supervisor, as a program intended to get high school drop-outs off the streets (where it was feared they were engaging in socially unproductive and criminal behavior), and back into the classroom, with a view to turning them into socially productive members of the community. During this time it was also referred to (by number) as the "2 to 4 Program", referring to its part-time hours of operation after the end of the regular school day. Shortly after I joined in 1985 it was expanded to a full-time program, brought within the regular school day, and renamed the "Core Program" - referring to its focus on the 'Core' areas of study: English, Math, Social Studies (formerly called 'History'), and Science. A year or so after this it was expanded again, to two full time programs, one focusing on the younger students (generally in their first two years of high school), and the other on the older students (generally in their last two). The teaching staff was doubled to 8, along with a part-time social worker.

Alternative programs in general have proliferated during the last half century or so in the American school system, driven in part by a 'rhetoric of crisis' strongly marked in its grey literature. This rhetoric of crisis has also driven the proliferation of new methods and types of teaching - one of the first I remember from my childhood was "New Math", back in the 1960's - today, the presentation of such new modes during regularly scheduled 'professional workshop days' and elsewhere has become a thriving business - my colleagues and I used to joke that we should come up with something "new and improved" with a fancy title (like: 'Teaching Aperture - a Way Out of the Box') and market it to schools nationwide in order to make more money than we were as simple teachers. -- Anyway, in addition to New Math, there has been the 'Whole Language' approach to reading and writing, 'hooked on phonics', 'back to basics', 'bell to bell teaching', 'learning to learn strategies', 'integrated teaching', 'cooperative learning', 'performance based education' -- to mention just a few of the grey buzz-words and attendant educational movements..

This rhetoric of crisis is, of course, not specific only to the last half century of American education; it is found in the literature of American advertising of this time, where products are continually being marketed as 'new and improved', 'better', 'bigger', etc. It also applies to the whole of the 20<sup>th</sup> century, from 'modernism' through 'deconstruction' to 'post-modernism' and beyond. Nor need we stop there, historically speaking: we can trace it back further to the grey pamphlets and posters of the revolutions

dating back to the 'age of enlightenment' (18<sup>th</sup> century). The concern that things are in crisis, breaking down (or in need of being broken down), and/or that things were better in the old days ('back to basics!'), that the proper balance between those four levels of social discourse (law, knowledge, business and industry) is in danger, is indeed co-extensive with civilization and/or politics (-- words from the Latin and Greek roots meaning 'city' respectively).

The clientele, the students, of such alternative programs as Horizons, are indeed generally in the midst of an identified crisis (a crisis of identification), as 'disaffected learners', 'at risk' of dropping out, burning out, acting out, and so on. For one reason or another they have chosen not to participate in school, or to participate in negative fashion - their grey literature production, instead of moving towards white or ideal literature, has turned black (non-existent) or 'red' (i.e. defined by the school and by law as a negative in need of erasure). Some are very bright, some are slow, with the majority falling somewhere in between. Many, though by no means all, come from poor or broken home situations. What *is* common to them all is that 'school', far from being seen as an 'alma mater' (nourishing mother), has become something threatening against which they rebel, a hated tyrant in their lives, an 'evil father' (malus pater).

To a lesser extent, these are all things that are found in the regular school as well. Schools are defined by law in America (government or legal GL discourse) as standing "in loco parentis" - in the place of the parents. They provide nourishment (in the form of education) and also discipline (in the form of rules and regulations to be followed). If the students take their nourishment and give it back properly (according to the rules) -- beginning with the learning of the alphabet, the 'ABC's', the letter - they are rewarded and praised as 'good' boys and girls, who are on their way to the 'ideal'. If they do not, they are punished or ostracized, as 'bad' boys and girls, on their way to nowhere.

Many students resist this encouraged and rewarded identification of themselves with the 'good' and end up counter-identifying with the 'bad' - a brief perusal of popular teen music reveals this very clearly. This ambiguity of identification can be found amongst schoolteachers and administrators as well (who were, of course, all students once themselves). I remember an anecdote the first supervisor/director of the Basic Studies program (precursor to Horizons) liked to tell. He (Charlie) was a science teacher in the regular school (a most excellent one, in my estimation) and had to go back to college for a supervisory degree before he took on the administrator role. In one of his 'supervision/administration' classes the professor read a set of rules and asked the class what they thought of them as rules for a high school. Charlie found the rules excessively harsh, as did one other member of the class. The rest of the class - most of them teachers working to become administrators - found the rules excellent, and thought it would be most excellent if all schools could adopt them. At this point the professor sprung his trap, and informed the class that the rules he had just read them were the rules for a federal penitentiary.

Having attended various 'alternative education' conferences over the years, I notice that alternative programs generally break down into two approaches or models. One sets up (in its GL) a very carefully defined set of rules and expectations, specific progress marks determined by standardized learning packets, and so on. The other is more relaxed, with flexible curricula that are constantly being retailored to motivate and meet the individual needs of the students presently attending the program. (Both these approaches, I should add, tend to work and meet with a certain success...) Horizons (and Basic Studies) followed the latter approach.

I mentioned "Charlie" as one of the first administrators of the program. This use of the first name (and avoidance of title) was a strategy that we carefully followed over the years. Students were encouraged to address the teachers by their first names, or by nicknames (which they were happy to invent/continue over the years). With regard to the "identification" ambiguity we found that this very much helped bypass the negative counter-identification problem. As teachers we thus stepped aside from the students' problems with authority figures - by stepping aside from our own presumed 'identification' - we entered the grey and stood next to them on the level, as it were.

Another aspect of this identification question, which the Horizons teachers found important to note and work with over the years, was the way in which individual students responded to gender. We would find that certain students would respond very differently to the male teachers than to the female teachers. (We generally worked to maintain a gender balance in both staff and student population.) Thus, for example, I as a male might find that I was having problems with a student (either male or female), which problems were shared by the other male teachers, but that the female teachers had no problems at all with the same student (and vice versa in the case of other students). In other words, for some students, it was the pater (father-figure) that was always seen as malus (evil), and the mater (mother-figure) as 'alma' (nourishing), while for others it was the mater that was mala and the pater that was almus. In some cases this was a direct reflection of the student's actual mother and father situation, in others it seemed to relate less to actual mother and father and more to the ways in which they had come to see other more generalized (i.e. social) authority figures in their lives. In any case we found that noting and dealing with this particular aspect of the identification problem made a marked difference in the success of the students.

As I indicated above, a major problem we as teachers had dealing with our alternative students was that they generally suffered from some degree of anomie - their grey literature production had faded to black: they would produce nothing; or their production, such as it was, had turned to 'red': graffiti on school walls and other grey acts generally perceived as negative. As an English teacher (of reading and writing, language and literature), there were three exercises I found worked well to break through this resistance on the part of the students, to overcome their 'writer's block'. Although the rest of the curriculum would change from year to year, these three exercises always found a place in my lesson plans at the beginning of the school year.

The first consisted of reading them a short passage from an author like Faulkner or James Joyce, written in stream of consciousness style, then turning them loose with a goodly number of blank pieces of paper. The goal was to see who could fill up the most pages, without regard to spelling or grammar, ordered sentences, etc. Most students responded well to this - interestingly, the most resistant could also be encouraged to participate, and would, though they held on to their resistance through the mechanism of repetition (lists of nouns, phrases repeated over and over...).

The second exercise was to provide them with a number of blank sheets of paper, have them hold one aside, and then on the others write as many observations as they could about that first blank sheet of paper. (The science teacher in the program would generally follow this up in his classes with slightly more structured lessons on the practice of observation.) The students tended to average about 50 discrete observations, with a bit of encouragement, in one 45-minute class period. (The all-time record was somewhere around 300, I think.)

The third exercise, and by far the most popular, was the so-called "story-slam". (I only discovered the name for it after several years, from another English teacher that joined the Horizons Program at that time, who used the same basic exercise as lesson himself.) I see it as a sort of primitive or basic grey literature generation program. Classes in Horizons averaged about 8 students each - for this exercise we would sit about a table (myself included - I also participated in the stream of consciousness and observation exercises). Each would spend about 3 or 4 minutes writing the beginning of a story, and at the end of the 3 or 4 minutes we would pass what we had written to the person on our left, who would pick up that story and continue it for the next 3 or 4. I always made it clear that there was to be no particular censorship involved - the students were free to write whatever they pleased. Names of the participants often figured strongly in the stories generated (myself included), with a good bit of rough give and take occasionally -- generally most took it all in with good humor -- it was, after all, an even playing field. From time to time I would perform an intervention (beyond what the other seven in the circle were writing), generally a brief private communication between myself and one of the students.

But here already we can see how, as soon as the students start producing again, another resistance beyond theirs threatens to arise: that of the school to aspects of the students' GL production. This suggests itself as the original resistance, against which theirs developed as an angry reflection. I side-stepped the problem of this return of the original resistance in the case of the story-slams by 1) participating within the group on a level ("keeping it real", as the phrase goes) and 2) functioning as 'expert' advisor to individual students rather than as 'master' (actual censor). Further evolutions of the students' GL, however, beyond the primitive and immediate story-slam, had to face this problem of resistance again on a higher level:

Along with a regular yearbook, once the program had enough computers and computer access, I would have the students work together to publish a newspaper or journal during the year. The first one was called Horizon Times, and ran for about 3 years, using Apple II's Newsroom program.

Like the story-slams, it was self-censored, if at all. (As the "expert", I would make recommendations on wit & lack of wit, but only go so far as observations on permitted and not permitted.) I can still remember the curiously ambivalent look on our administrator's face when he saw and began to read the first copy. He (Mr. G in this case, not Charlie, who had been forced out of the position earlier) was enamored of the notion of "project teaching" and thrilled to have an example of it from the program he was administrating. At about the same time as he was laughing at the somewhat intimate 'Advice Column' in the Horizon Times, a look of despair started to descend. He did not need to tell me - at that moment I could read his mind. When he first saw the product he immediately wanted to show it to the Principal and even the Board of Education, as an example of what Horizons was accomplishing. Reading it he realized that he could not do that in any sort of official capacity: amusing and as well done as it might be, the content had the potential for upsetting certain people and drawing censure. -- So we did not show it off officially, but rather let it flourish half-hidden in the grey. (The student who had written the advice column actually did bring a copy of it to the Principal of the school - unofficially, as it were; from what she told me it seems the principal's reaction was much the same as Mr. G's.) We usually made extra copies for students outside the Horizon's Program in the regular school where it and the later New Horizons Literary Review (done on Windows based machines) were popular. -- Neither made it to an official presentation at the Board of Education, though a few selected members of the board did get copies from time to time.

-- A theoretical detour is in order here, to elucidate this problematic of resistance in and to grey literature production. Let us take a closer look at the Lacanian algorithm of the four discourses mentioned above, and at how it represents the four discourses of grey literature. (I use here the model as it has been developed by Dr Theaux. The pictures you will see come from some of his webbed-material, where they appear as a single rotating .gif file.

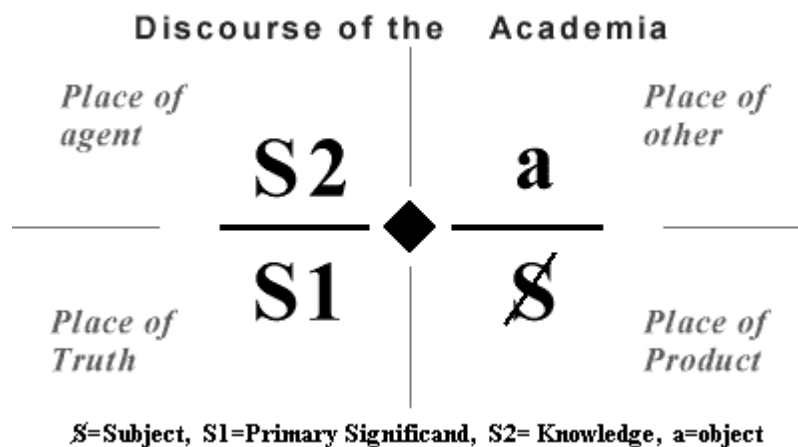
[www.akhnaton.net/dna/sub01/1999/20030823155300\\_knwan\\_fr.htm](http://www.akhnaton.net/dna/sub01/1999/20030823155300_knwan_fr.htm)

These depictions show how 4 terms or images (/S/, S1, S2, and a) rotate about 4 places (agent, other, product and truth) to produce four discourses: academia (the teachers), government (the board of education), industry (the students), and business (the administrators).

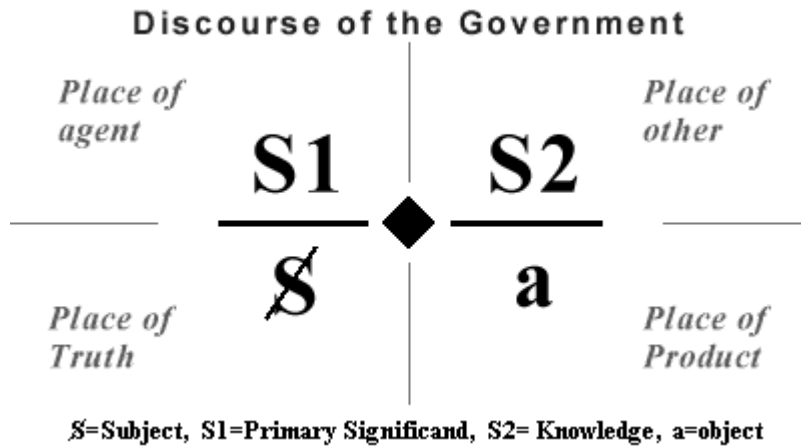
'S1' stands for the primary significand, the written letter, the spell of a word on a page. 'S2' is the second significand, whose necessary interaction with the first makes 'reading' possible. Thus it stands for knowledge, or knowing. The barred '/S/' stands for the individual Subject. The 'a' (or aobject) stands for all objective other ('a' objects), and by extension the environment in toto. These 4 represent the basic necessities for a discourse: a Subject, an Object (<>environment), and the linguistic doubling with which a discourse proceeds (S1, S2...)

The 'Place of agent' indicates what it is that drives the action of the discourse. The 'Place of other' indicates what is reflected in the discourse. The 'Place of Product' indicates what is produced by the discourse -- and thus at the same time reveals the cause of the discourse. The 'Place of Truth' indicates what is seen as essential in the action of the discourse.

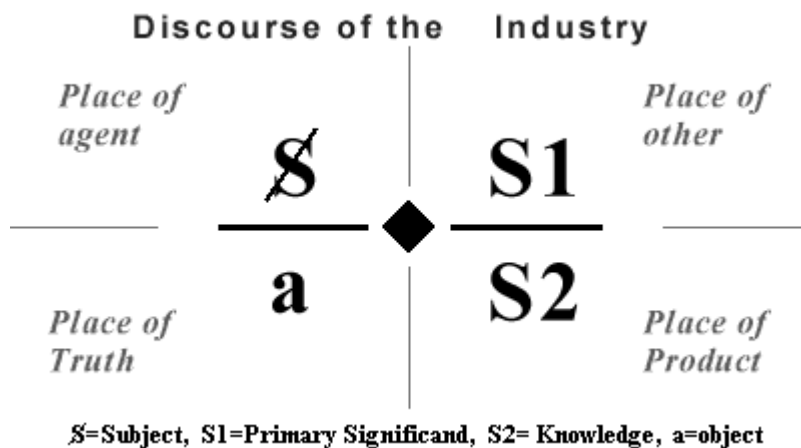
Let us now begin the rotation of these images on these places, and show how the 4 permutations, in this model, generate the four discourses of grey literature.



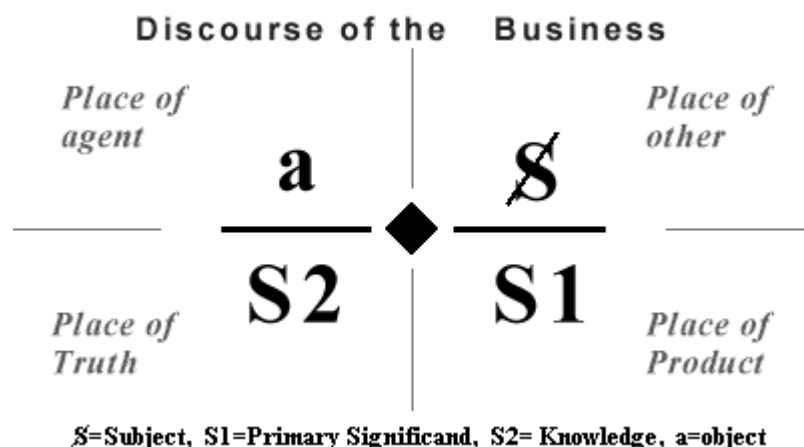
In the discourse of academia the Subject (the individual student or teacher) holds the place of the product: the individual subject is what is educated in the discourse of academia. This discourse addresses itself to the production of the Subject. Knowledge, S2, holds the place of agent: knowledge acts to produce the subject, and for the sake of the subject. The written letter, S1, holds the place of Truth: the written letter determines the truth of the discourse. The Place of other is held by the 'a', the environment, which the written letter maps in 'truth' in the discourse.



In the discourse of the government the environment, all that which contains subject, letter, and knowing, holds the Place of product. This discourse addresses itself to the production of the environment. The Place of agent is held by the written letter: the letter of the law, in this case. The Law acts to produce the environment, and for the sake of the environment. The Subject holds the place of truth: the Subject determines the Truth. The place of other is held by knowledge, which the Subject maps in truth in the discourse.



In the discourse of industry knowledge holds the place of product. This discourse addresses itself to the production of knowledge, of “how to”. The place of agent is held by the Subject. The Subject acts to produce the knowledge, and for the sake of the knowledge. The place of Truth is held by the environment: the environment determines the truth. The place of other is held by the written letter, which the environment maps in truth in the discourse.



In the discourse of the Business the Place of Product is held by the written letter. This discourse addresses itself to the production of the written letter, (in this case money-notes as capital). The Place of Agent is held by the environment. The environment acts to produce the written letter (money/capital), and for the sake of the letter. The Place of Truth is held by knowledge: knowledge determines the truth. The Place of other is held by the Subject, which the knowledge maps in truth in the discourse.

The discourse of government, by means of the letter (of the law), produces an environment for the four discourses; that of industry, by means of the subject, produces a knowledge; and that of business, by means of the environment, produces a letter (money). The discourse of academia, or education, by means of knowledge, produces a subject for the 4 discourses, either as teacher (servant of knowledge), or as leader (servant of the Law, the letter), or as business-person (servant of environment), or as worker/laborer (servant of Subject). These four subjects define four basic social roles in civilization's discourse - and it is here that we can begin to trace the problematic of resistance.

The students I dealt with resisted identification (some actively and some passively) with any of these four roles - most in response to an earlier resistance to (and rejection of) their own first attempts at grey literature production.

As I indicated earlier, there are two basic ways in which alternative education programs approach such students: one tightly organized with explicitly defined rules and standardized learning packets, and the other more relaxed, with flexible curricula tailored to the students. (Sort of like McDonalds and Burger King styles, I guess, "we do it all for you" and "have it your way".) The first approach offers students a chance to return to their primal educational scene and to get it right (redeem it): this is presented as an attainable ideal by this approach, its pursuit reinforced by careful structure and regulation to effect behavior modification and overcome the resistance. The second approach also offers return to the primal educational scenes, but focuses instead on rediscovering the desire the students lost (or displaced) in that first resistance, and then encouraging it to flow back into the rotation of the four civilized discourses (as opposed to being frozen in or against any given one).

This latter approach aims to discover something hidden (the repressed desire), and points to a further theoretical consideration that applies to all literature and literary discourse. Literature consists of significands - letters, marks on a page, which refer to a signified. The act of reading consists of translating these significands into what is signified by them, according to given conventions. Thus we translate "table" into a flat surface about waist height on supports, and not into a small furry animal with wings. -- However, there is something that hides itself from this act of translation: namely whatever else, beyond the conventional meaning, the shape and form of the significand itself can suggest to our eyes.

To illustrate this more fully, let us take as an example the phenomenon of graffiti, which I mentioned earlier. It is generally seen (by teachers, school administrators and the law) as something negative, an act of defacement - of public buildings, restrooms, and so on, up to a 'defacement' of civilization in general. For the moment, however, let us put our resistance to it aside, and look at it as a sort of prime example of grey literature, of unregistered textuality, and analyze it as such.

The most common form of the graffiti is the 'tag', or name. As such it indeed functions as a signifier for something signified. But in the distortion of the letters, and in its placement - on a public building, or on the body as tattoo -- it emphasizes its nature as sign; it stresses something that escapes the translation to what it conventionally signifies.

An extreme example here: some seven or eight years ago tattoo parlors started putting up charts which gave a set of correspondences between the English alphabet and Chinese ideograms. A number of my

students used that chart to have their names tattooed on their bodies in Chinese. Well, it was not long before people started pointing out that this set of correspondences (which may have originally popped up somewhere on the internet), was entirely arbitrary, random, and with no actual relationship to the Chinese characters. Now, one would expect the students with such tattoos to be embarrassed somehow, but that did not seem to be the case. They tended to meet criticism not with embarrassment, but rather with a slightly defensive puzzlement tinged with contempt. It did not really matter to them that the relationship between the English and 'Chinese' spelling of their name was arbitrary - they saw the relationship between signifier and signified as arbitrary - conventional - to begin with -- and in the case of the Chinese 'letters' that conventional/arbitrary link was established by the charts shared by those getting such tattoos, even if nowhere else. What was really at issue for those getting the tattoo was its function as sign, for what it said beyond what it signified.

From the perspective of the various graffiti-artist 'taggers' I came to know over the years I found that they themselves saw their activity generally as a desirable/joyful artistic and literary, letterary, production: the standard letters, the ABC's one learns in the first years of school, were transformed by their work into "signs". They turned the standard (and to be repeated) letters into something new, without entirely losing the recognition that comes with repetition. Graffiti was for them an affirmation of the production of the letter - a return to the primal scene of literature, something they approached with a combination of awe and desire.

Graffiti has worked its way into the letterature of popular culture as well - note the tendency in the last decade or so towards "corporate" tribalism: i.e. logoed baseball caps and sweatshirts, t-shirts, and so on, with the names of manufacturers like nike adidas reebok fubu etc - done up with graffiti style lettering. (Not to mention Gucci and Effendi handbags...)

Interestingly, the alphabet as we know it may have had its origins in graffiti. Fairly recent discoveries at Wadi-en-Hod just north of Thebes/Luxor in Egypt reveal the names of various Semites who were working there around 1800+BCE, carved into the cliff-sides, in a script abstracted and simplified from Egyptian hieroglyphics, in a form which shows a clear connection to the proto-sinaitic and proto-canaanite scripts which later evolved into Hebrew and Phoenician, passing from there to the Greeks to become the basis for the western 'alpha-bet(a). (<<http://www.library.cornell.edu/colldev/mideast/alphorg.htm>>)

What we can see from this analysis of graffiti is that language, literature, the letter, consists not only of a binary relationship between a signifier and a signified, but at the same time supports a ternary code operation, whereby a third term lies hidden in the encryption of the letter - a term that is generally repressed and forgotten in conventional discourse, (i.e. it remains in the Unconscious), but which activities like graffiti - and related art forms such as calligraphy - not to mention poetry and song - bring to the fore.

There is something else that seems to bring this code operation to the fore, or at least inspires the same sort of desire/joy exhibited by graffiti artists, and that is the cybernetic apparatus itself. - And thus, with these theoretical considerations in mind, I turn finally to the effects of the introduction of computers into the school system.

America tends to be technophilic. Not long after the appearance of the first personal computers, they began to be introduced into the school systems. Grant moneys hastened this process. In Branford, when I began teaching in 1985, Commodore 64 machines were being used - some in various classrooms and the majority in specialized rooms (business, reading). These gave way to the Apple II machines and then a variety of IBM clones. In the 90's record keeping (grading etc) was transferred from paper files to networked computers. Towards the end of the 90's Branford received a \$1.4 million grant to hook up to the internet, with media centers in the library and faculty room and computers in almost every classroom.

For all this, a certain resistance to the use of computers manifested itself from the beginning, on the part of administrators and teachers. Youth have bonded strongly with computers - especially the 'game' machines - and in the early phases of the cybernetic introduction into the schools teachers would worry about students bringing in games to play on the computers (rather than use them for specific 'educational' purposes). Later, with the internet, came the additional worry that they would be using the internet to visit non-approved sites - for which pornography has become the prime emblem. Thus school system intranet 'firewalls' were introduced that also kept students from researching things like 'breast' cancer (keyword forbidden).

These firewalls, however, proved to be vulnerable to student hacker-style manipulation - once when I mentioned it was too bad the firewall 'safeguards' also blocked legitimate research, a student gleefully informed me that all I had to do was put an extra period after any given URL to get through to any site. Well, firewall administrators figured that one out and 'fixed' it the next year, whereupon students, of course, applied themselves to figuring other ways around it.

The use of computers to write generally increased student production. Many who could hardly be convinced to write on paper with pen would happily sit down at a computer to write - with a marked tendency to increase the font size to fill up more pages. Similarly with research: students who tended to



withdraw into a passive shell when taken to the library often became some of the most active when set before a computer to browse the internet. There seems to be some magick in the machine that reawakens desire in many students. Some of this can perhaps be explained by libidinal attachments created playing video games on machines carrying over to the other - though I suspect it is ultimately the hidden (but obvious) magick of the cyber-code and resulting play of signs that charms the "disaffected" learner back in to desire to participate in both cases.

This renewed and increased production on the part of students engaged with the cyber-interface has raised other problems (more resistance-formations): common to many schools now is a renewed concern with 'plagiarism', proper citation, and right to copy generally. Many of the students active on the internet have a rather cavalier attitude to such things. If one can simply copy/paste and share the information that way, why go to all the extra bother of paraphrasing? Marshall McLuhan's predictions of the global electronic village returning to medieval (pre-print) attitudes to authorship seem to be coming true here. I much doubt the (fairly recent) notion of copy-right will survive, ultimately.

The internet, with its e-mail, omnipresent chat-groups, forums, instant messaging, web-sites, has proved to be very attractive even (and/or especially?) to the recalcitrant learners and writers I had as students - and is changing the nature of writing as it increases its production. I remember the first time I walked through the media center room in the high school, with a dozen computers or so, in the first year of internet presence. I noticed all were in use, though only one was being used to browse the web/do research and/or write - the others were all being used to play internet accessible games (and cheap ones at that!). Over the next year and a half these percentages changed radically: at the end of that time, when I would walk through the media center, on average only one computer was being used to play a game, one still being used to research/browse -- with the other ten being used by students to instant message and write e-mails!

By the end of my tenure, another phenomenon was growing: that of students who were putting up their own web-sites, their own grey literature, their own networked information, independent of institution and school (and thus free from resistance and censorship). They were almost always most happy to show them off and share them with others.

The impact of this at all levels - it is not just students who are putting up web-sites and networking across the AI interface of the world wide web -- on the evolution of the four discourses of civilization (academia, government, business, industry) will be great - most especially, I suspect, on the institutions that presently maintain them.

Branford, CT, USA, 11/2003

---

## Author Information

### Dr. Harold von Hofe

Horizons

32 Medley Lane

Branford, Connecticut 06405

USA

[h169@sbcglobal.net](mailto:h169@sbcglobal.net)

## **Body and Literature:** Study on the Grey Literature of a health service starting from a reflection on health

Dr William Théaux  
psychiatrist - psychoanalyst

### Abstract

Existence of Grey Literature can be followed, as primal ego is followed by Psychoanalysis after considering an Unconscious at its source. In case of Literature, Cybernetics is found in place of such source; this has consequences in regard with the human bodies that are living along with this Literature.

### **Initial Perspectives**

This study intends to examine the way in which an institution manages its Grey Literature at the current time, so remarkable for its development of Information Technology. This technology is endowed with such an interactivity nowadays that it deserves the general designation of cybernetics (networks, digitalization, interactivity looking towards an effective artificial intelligence). The institution I examine is old and well organized. It is an establishment of the Red Cross which is at some distance from the decision-making centers, in the countryside, and is barely computerized. It has a good tradition of files, memoranda, advertisements, postings, tables of activities etc... With cybernetics, this tradition ages like red tape, and we are working towards an identification of a new function, known as Grey Literature.

I begin by cataloguing and detailing three elements: the **object** of this study, then its **goal**, and thirdly its **conditions**.

First the **object** of this observation. It concerns itself with the identification of GL in a Medico-Pedagogical Institute. This is an institution which brings together care-givers, teachers, and users, the latter mentally (or socially) handicapped children.

The **goal** of my activity, in this place and in my capacity as doctor, is that of *health*. I understand by this term in particular the health of the members, and in general that of the institution. In this regard I must mention my earlier studies and researches, which focus on the health of social bodies. It is thus in relation to an integral health and sanity that I perform an analysis which observes, in the reciprocity of the individual and the collective, the function of grey literature.

Thirdly the **conditions** of my work are characterized by my medical specialty, which is that of psychiatry - that is to say that in my engagement for the health of the bodies, whether social or individual, I necessarily include the function of *psychism*. Moreover, within that specialization, focusing on psychoanalysis, I consider especially the image of the body, which is formed, in human reflection, relying here on modern conceptions which attribute to this image - which one calls *imago* - an extreme dependence on language. Now, according to these modern

conceptions, language, thus considered, is most strongly structured by the Letter - which is to say, when all is said and done, by literature.

### Perspectives of analysis

We may thus gather together the conditions and terms of my observation:

That which one calls Literature - or *textuality* - in the institution which I am analyzing, presents itself in a manner intimately coordinated with the psychism of its members, it is in the dependence between psychism and the body that I am required to maintain my deductions. I can thus set the departure point of my work and project with the linkage Literature>psychism>body and the following proposition: ***the activity of psychic care in an institution provides the occasion of examining or verifying the link between literature and the identification of the body.***

There remains to be seen how this link between literature and the body reveals itself in the light of technological progress. From the perspective of the body we know, for example, that in our times its definition refines itself considerably with an identification of an *individual body* according to the criteria of a genetic code. On this basis, which lends itself to a comparison of chromosomes with a sequence of letters, ecology often invites us to make a connection between the biological codes and those of the environmental niches, which the *social order* strives to reproduce in symbolic context. It is on this level that we find literature, properly speaking, which, in its turn, returns to the level of the code under the effects of its cybernetic treatment. We know now that it is the emergence, treatment and objectification of Grey Literature that indicates an aptitude on the part of literature to respond to a logic of code.

Therefore, reconsidering at present the Psyche, because we have isolated the code, that is to say the regime which brings into direct relation individual body, social organization and literature, we are able to examine the framework of confinement of this triple combination, of *person, group* and *textuality*, which constitutes this institution called *Faidoli*.

### Mutation of perspectives

This establishment was principally an *internat* - an in-patient service -- for many years. It thus represented a closed space where the children lived during the week. One often thinks of such places as a second family, but above all symbolically as a mother, holding her children in her bosom, in her womb. It was during this year 2003 that an important transformation affected this *internat*; in phase with the effects of the modern world, an essential part of *Faidoli* transformed itself into an out-patient *externat*.

It is in the midst of this transformation from an *internat* into a Service of Care in the Home (called SESSAD) that I present my study of the identification of Grey Literature. Now, instead of living, eating and sleeping in the institution the children reside normally with their families. And, inversely, it is the educators and care-givers who visit them, in their homes and schools. The children have separated from the mother, or matrix, who used to contain them. The connections between the institution and its members have psychically inverted themselves, with the matrix now on the outside -- similar to when children are born and the mother comes to them.

Such a reversal can be expected to affect the literature in this SESSAD. According to the strong link we see between the human body and the symbolism as the individual and his world - through the media of psychism as Mother, environment, language - a large set of significant combinations will provide occasion for translation into its textuality. We call it the **phantasm**, and further, it is by the distinction of its Grey Literature that one sees how this textuality becomes a complete part of the identification of the body.

Yet we must be cautious; it is always necessary to measure the risks and difficulties of such a view of the human identity. If cybernetics nimbly invites us to equate body and text -- as the image virtual reality suggests, we sense at the same time that this leap into a pure world of code may be only a dream - and since, practically speaking, this establishment is only very

minimally computerized, we sadly do not have an abundance of indicators to objectify in terms of networks, programs or procedures. Hence we cannot measure how closely an improved health relates to a functionalized Grey Literature.

But on the other hand the necessary prudence and indigence of the circumstance present a situation suitable for the release of theoretical propositions.

### What is Literature from Psychoanalytic point of view

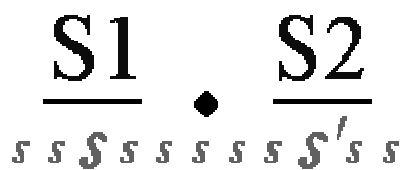
The topology of the framework and its bodies having been identified, we can now bring to it the final element. We have to start with a question: how does GL integrate itself into a definition of literature in general? It is a current question; Dominic Farace reminded us of it recently à propos of the grey literature in the networks, and the difficulty it has leaving them. On this point psychoanalysis can bring its brightest lights to bear. In effect, the analysis of psychism since Freud has taken the direction of a theorization that combines it more and more with linguistics. By hypothesizing an Unconscious it suggests a **conception that the Cartesian 'ego' of the human person results from a hypnosis whose roots are found in the Letter and alphabetization**. I recall two/three notions: the historical source of psychoanalysis is the phenomenon of hypnosis, which Freud decided not to practice, but to study; it was from this moment that he suspected that it was a phenomenon essentially linked to language; and following his observation of dreams, jokes and lapses he identified this link in the effects which the Letter, in practice reading and writing, imposes on language.

Since the 'ego' which forms civilization is a reflection of the effects of the Letter, the work of a psychoanalyst assimilates itself to that of a linguist. Hence, in reciprocal relation, psychoanalysis brings its contribution to the general linguistics of literature, and in its turn the theory of the nature of literature enlightens psychoanalysis. Here in the form of an aphorism, is the basis of this reciprocal relation:

**Just as the *psychism* reveals to *linguistic analysis* that it is composed of an *Unconscious, literature*, in its encounter with *cybernetics*, reveals that it comprises a *Grey Literature*.**

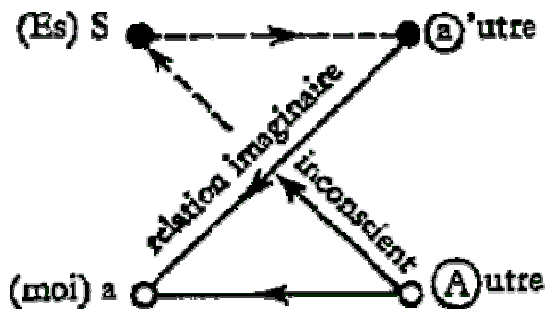
But the usage of aphorisms comprises a limit too - just as language which, the more clear it becomes, the more ambiguous it is; we hardly know how to doubt it, we who suspect that the perfect semblance of an achieved literature, edited, proprietarized and commercialized, hides, in some sort of inversely proportional manner, a literature grey and pregnant with an essential truth, possibly repressed. Thanks in particular to psychoanalysis, one understands that the dazzle of the letter in the meaning conforms to the fact that one cannot, by holding oneself to the letter alone, distinguish between lie and truth. To break this duplicity, which appears also as the sort of trickery between literature and propaganda, scientists know the solution under the name of algebra.

The founder of Linguistics, Ferdinand de Saussure, already formulated the rudiments of this algebra in his science. His description of the sliding of the signifiant over the signified is well known:



Psychoanalysis has continued to develop these formulas by allying them to cybernetics. Thus it presently proposes the following formula to clarify the definition of literature - which constitutes the fundamental diagram. It is the schema L ('L' for Lacan, if you like, as it was he who introduced it to Freudism) - it is also called formula Z - which describes how the Saussurian

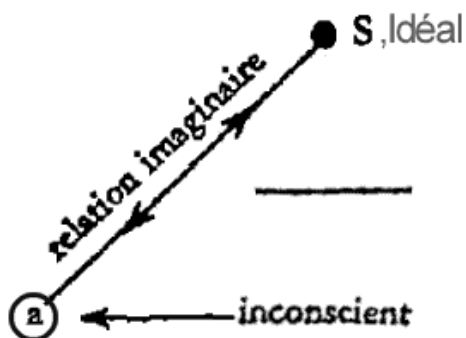
significand hooks the signified meaning (lower level), in spite of its slide (upper level) which seems structurally to detach it indefinitely.



Schema L.1: This formula exposes the narcissistic relation (see: *relation imaginaire* - imaginary relation) of the human relation through which a structure (see: *inconscient* - Unconscious) tends to signify itself (A), by alluding to a Subject (Es) whose reflection (a)' it would be.

In fact the schema L is essentially inexact because - since it deals with a formulation of cybernetics: it is dynamic and must change when writing itself (this writing takes the form of the four discourses of which Mr. von Hofe spoke). One remembers the classic example of cybernetics: *two machines which each react to the reaction of the other and in this regard realize a perpetual instability*. This imprecise but fertile mislaying of the action in its effects expresses itself mathematically in the famous figure of the *catalogue of catalogues*. For the mathematician it may have to do with the cipher zero and, for the librarian, with the grey literature in the inventory of his library - adding for the latter the practical anguish, when he becomes aware that the cipher has materialized in his library with the introduction of cybernetics. The schema L proposes itself to describe this vital knot of irrepressible instability which literature supports, and how cybernetics reveals it:

schema L.2



In psychology, the schema L shows that as soon as the imaginary relation between counterparts *a* and *a'* rises to awareness - which is to say as soon as it assumes that the subjects which it potentializes are dependents of an Unconscious - then this primitively narcissistic and transitive relation polarizes itself in two terms of which one constitutes an 'ideal subject' (schema L.2 : S, I) and the other the most objective object (: O, a).

Nevertheless this object, if it is perfect as the schema L.2 analysis of the libido attests, is to include the desire one calls *by-product of the Unconscious*; in order to constitute that which one calls the *ego*, it includes itself

as an object of drive (*a*) in an envelope, body, sphincter, that is to say an objective form (schema L.2 : circle around (*a*)).

This formula applies to literature as well. When it lets two states be seen, we are in the habit of calling them *grey* and *white*. According to the schema L this opposition composes itself along the line which is called imaginary relation, where White Literature is more clearly understood as *Ideal Literature*.



### schema L.3

The *Grey Literature* which distinguishes itself here also reveals its nature more clearly: after the manner of the psychoanalytic object, in the appearance of multi forms (drafts, e-mail, tickets, bills, plans, which is to say the classical supports of GL) this literature conceals the ultimate object (a) of a *drive* of its enterprise.

Thirdly, this formula L, of the Letter and/or of literature, indicates also the place and the function of the cybernetic apparatus (LAPAREIL) which uploads, as the Unconscious does with its own cause, the embodiment of said GL

### Study of the distinction of GL in the experience of SESSAD

We can now return to the institution SESSAD, to test this formulation of literature in a concrete case. By the side of the old and large internat, SESSAD provides care to twenty children. The team that provides care and education consists of twenty persons. It was launched in September 2003, only a few months ago. In these days of December 2003 we can only see the inherent instability of beginnings, which masks the cybernetic instability inherent in its embodiment of GL. My testimony and my interpretation respond to the following two levels I can observe: a first level which is start and facade, and is created by the resistance which the analysis of psychism necessarily arouses. The second level which also always offers itself, beyond this resistance, is that of theory, in the present case of the nature of GL.

Let us begin with the resistance aroused by the inception of change; in its opening period, SESSAD had counted on receiving a computer program to manage its GL. After many months and in an inexplicable manner, the organization which was to deliver it, has addressed only the documentation and mode of use, but not the diskettes or the program. Secondly, in the very days of its opening, the institute Faidoli repeated a general failure to equip itself with a computer program for outside communication, Internet, or other mode of communication between SESSAD and its supervisors.

The character of an entrenched citadel, in a period comparable to a flood where the prophet Noah gives the lead to his daughter Knowing, testifies to the resistance with the fact which psychoanalysis has already made it possible to bring to light:

It is known indeed that the tension between individual and collective psychology has held psychoanalysis, faced with the barrier of the transference, in immobility. The analysis of the neurosis of transference concludes with this parabola: *the discovery of the Unconscious presents the risk of denaturing the ego in a Collective Unconscious - against which the parade veers in the other direction of confinement in the transference, which gives place to the death instinct. If things are comparable, literature, discovering cybernetics, risks denaturing the private life and the rights of the person - against which the false parade into isolation, with only an intranet, risks turning the grey literature to black, as transference turns to the profit of Thanatos.*

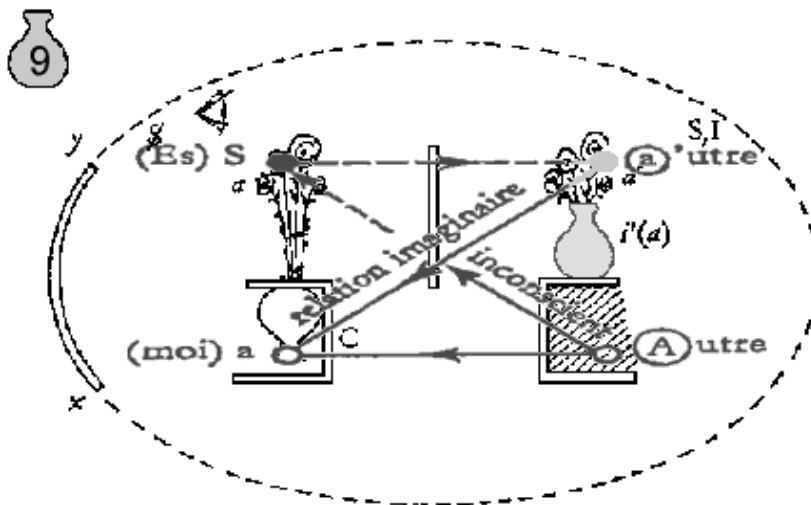
Since this is not a course on psychoanalysis I cannot detail the much debated logic of the Transference and its certification of failure. For this reason I exposed it in terms of a parabola. I exposed it in any case because there would be no way to have a theorization of GL without the notion of the barrier of resistance. It is only then - as when one, by breaking the sound barrier, separates an object from its noise - that one can, in the silence that lends itself to study, analyze the theoretical *corpus* of GL.

### The human body in the institution - held in grey literature

One can recognize the nature of GL by that which motivates attraction to or fear of cybernetics. In psychoanalytic terms one asks: what is the object of desire which corresponds to cybernetics? From the perspective of Literature, a first opinion comes naturally in response to the question: *the desirable object of literature is knowledge*. Is it for this reason that one

finds resistance to the grey form of this literature? One can understand how a knowledge in rough draft, uncertain or even erroneous, poorly controlled or deceptive, sets in motion an affect of rejection, for fear of losing or harming this ideal *knowledge*. However, this first opinion is put in question again when cybernetics informs us that, on the contrary, its goal is to effect an ordering of this confused knowledge. We must thus admit that it is some other motive than ignorance which motivates contempt and rejection of Grey Literature. It is precisely in the formulas of cybernetics that we are able to find this hidden motive:

When we take up the series of *L-schemas*, 1, 2 and 3 above, and evaluate this series with regard to its psychic formulation - otherwise called *optic* -- we find again, at the place of GL, that which corresponds to the most biological basis of psychism.



**Model O: extracted from an animated picture - here : general psycholgy phase**

The Optic Model shows in the form of vase/containing and the flowers/contained the psychic imagoes and their physical correspondents in a mirror game analogous to Schema L.

We have already remarked how, faced with an Ideal literature (formerly called *white*), *grey* literature presents itself as *materia prima*, veiling in the envelope or form of temporary and non-commercial media, a rough object (a), identified in psychism as object of desire. Now all the trivial sexual histories of the human person remind us of the resistance as much as the necessity of recognizing that completed envelope with its object as the other's body. So many neuroses and perversions oblige us to ask if we know exactly what we wish to say when we are speaking of our body. The premises of Virtual Reality have prepared us to consider the type of form with which we identify ourselves - up to the extravagance of clones, and genetics to finish, whose finger points to the heart of the enigma that constitutes the body. At this heart of all perspectives, psychoanalysis shows that the body delegates its form to an *ideal ego* - which, we believe, depends on an environment. - "Bathe a body long enough in water", Darwin will have said, "and it will become fish." The new sciences thus incite us to make the effort of admitting that the body they have undertaken to explore has proven itself unknown, though probably localizable and perceptible by its formulas.

As it happens, in the place where GL falls in the examination of Schema L (schema L3 bottom left), Model O shows, by means of the allegory of a vase, the real body (Model O bottom left), which the imaginary one, far from the contingencies of code, idealizes (Model O top right). Consequently, if psychoanalysis is exact on this point: GL, in its opposition to ideal literature, corresponds to the real body of the human being.

Obviously, we could initially reassure ourselves, saying this is just theory. And secondly we could say that it is because my argument develops itself in the context of health, which leads necessarily to a taking into account of the body which is the object of an institute of care. But we can also tell ourselves that it would be an elegant solution which would explain the

resistance, otherwise enigmatic to GL. Because finally we come - if it does not concern only care - to ask ourselves if GL does not constitute the political body of the human being - understanding why politics - one sees this in the treatment of the masses in history - makes so little of the human being's biological body.

Perhaps we should not fear to go that far, we who are devoted to the identification of GL. Because even if we pretend to put it in relief within Enterprise and the world of Commerce, we will not forget that in its claim to sustain the political body, GL finds itself at the source of the Code - and thus in this corner, "bottom left", in a position of Truth to ethically direct the commerce and enterprise of a politics of an ecological body. It was important to me to evoke with you these very ambitious, but also very dynamic, perspectives for GL, although it is possible that at this point, the artificial intelligence alone will serve to further my purpose tomorrow.

#### NOTES

- 1) Schemas are from Lacan's Teachings, *Ecrits & Seminar*.
- 2) author's previous essay on *GL & Unconscious* can be found on the Internet  
French : <http://www.dnafoundation.com/regis/cd228/ijgl/raclig.htm>  
English : <http://www.dnafoundation.com/regis/cd228/ijgl/raceng.htm>

---

#### Author Information

**Dr. William Théaux**  
psychiatrist - psychoanalyst



2bis Fbrg St. Jean  
43000 Le Puy en Velay  
France

wtheaux@club-internet.fr

## The ISFOL Specialised Documentation Centre: an important tool for operational research

Isabella Pitoni and Diana Macri  
ISFOL

This paper aims to present the ISFOL Specialised Documentation Centre (CDS) and its activity in the production and management of grey literature in the field of vocational training and labour policies.

There are many services offered to researchers and operators related to GL:

- The database of all ISFOL documentation (which includes a large number of GL documents)
- The LOGOS database, including all documentation (published and grey) collected by the SDC multimedia library
- The implementation of the ISFOL GL in the SIGLE database
- Co-ordinating the publishing of periodicals on documentation, carrying out specific projects for the publishing of handbooks, guides and thematic catalogues
- Promotion of ISFOL/SDC documentation products and services on the Internet.
- Managing a specific documentation system aimed at promoting information on EU policies with particular reference to FSE.

We will demonstrate how the SDC has an important role in the dissemination of information and documentation (mainly grey literature) on vocational training and labour policies and on the European Social Fund for which ISFOL is the institutional body for the Ministry of labour and social policies in Italy.

### Introduction

With this paper we would like to present the ISFOL Specialised documentation centre GL services and documentary products in the context of ISFOL institutional activities and in the framework of European union policies and strategies for the European social fund.

We will show how they represent a fundamental tool for research in the field of vocational training and labour policies in Italy and on the European social fund.

Grey literature and in particular public sector grey literature is today a preferential means for the diffusion of information and communication on specific topics and ISFOL recognizes its key importance.

Especially in the context of the development of new technologies GL has found new and faster means of diffusion and improved its visibility and efficiency. At the same time next to the numerous traditional GL documents new forms of GL are developing in the Internet.

ISFOL through its SDC is taking fully advantage of these new opportunities and is active in exploiting the Internet for improving the visibility of its GL and for providing new services.

### The ISFOL Specialised Documentation Centre

A brief introduction of the ISFOL, Institute for the Development of Vocational Training for Workers, of which the Specialised documentation centre is a project, is necessary.

ISFOL created in 1973 is a state-owned scientific research institution engaged in promoting and developing vocational training and labour policies in Italy. It carries out its activities in collaboration not only with the Ministry of Labour and social policies, but also with Regional authorities, Social partners, State departments, the European Union and several international organisations.

Between its institutional tasks we find the promotion of study, research, documentation and information activities and the exploitation, diffusion and transfer of the results of its activities.

Over the years, the institutional objectives of ISFOL have developed in harmony with the social, political and economic changes taking place in the country.

Since 1995 the Ministry of labour has requested ISFOL to set up the Technical Assistance Structure (TAS) for the European Social Fund under the European Social Fund 1994-1999 programming. The Structure's task is to provide support and assistance to the Ministry of Labour and Regional Authorities in managing and implementing actions co-funded by the ESF and envisaged in the Multi-regional and Regional Operational Programmes for Objectives 1, 3 and 4. ISFOL today operates actively within the framework of the European Union strategies for the European social fund.

The Specialised Documentation Centre (SDC) is one of ISFOL's projects set up in 1997. It is a Centre for specialised multimedia documentation on themes for which ISFOL is institutionally responsible and is intended to provide theoretical as well as technical and operational documentation answering the needs and interests of national training-system operators. Specifically, it produces and distributes innovative and integrated documentation products and services to support social and economic research and the design and management of training activities.

The SDC also operates within ISFOL's Structure for Technical assistance to the Ministry of Labour and Social Policies UCOFPL (Central Office for Vocational Training and Guidance) and it receives funding for these activities from the European Social Fund.

The project focuses on the following sectors of activity:

- Organisation and maximisation of the body of documentation produced by ISFOL
- Research for the creation of models and tools to transfer technical and scientific skills and know-how
- Design and dissemination of innovative documentation products and services
- Technical assistance to the Ministry of Labour and Social Security - UCOFPL - Action 3.3. Specialised Documentation
- Structure for Technical Assistance to the MLSP-UCOFPL for the CEDEFOP Community Programme "Study Visits for Vocational-Training Experts".

In the context of these activities the Specialised documentation centre is very much involved in the management and distribution of documentation, in most cases grey, produced by all sectors of ISFOL. At the same time it also produces grey literature itself and is very active in its distribution and dissemination.

The SDC provides a wide range of documentation services for internal and external users, especially ISFOL's institutional players such as the Ministry of Labour and Social Policies, the Regional Authorities, bodies representing the social partners and managers/users of the training system.

### **GL Services provided by ISFOL Specialised Documentation Centre**

Grey documentation, intended as the one produced by organisations in the context of their institutional activities, has always been characterised by its fast distribution through informal channels, by its poor and basic means of production and by the difficulty in finding it, or even knowing it exists. Today the development of information and communication technologies and the worldwide achievement of the Internet has meant important changes for this kind of literature. Thanks to information technologies issuing organisations have new and more sophisticated means of production of GL and new tools for managing it in effective ways. The Internet is at the same time a rich information and documentation source and a new and fast communication channel.

New technologies have given organisations producing GL new opportunities for managing not only the production process but also the control, cataloguing, indexing and distribution of GL. All these activities are normally carried out through a documentation centre/special library that contributes to the achievement of the mission of the institution to which it belongs.

In the case of public organisations the importance of managing and providing access to the documentation produced is even higher. In the context of the development of new technology,

new public policies (at European and national level) has arisen inspired by principles of accessibility, transparency and openness.

The ISFOL SDC is a good example of how a public research organisation can, through a specialised sector, efficiently manage all aspects of its documentation life cycle and effectively provide access to information to its users.

We would like to present the documentation services offered by the SDC in more detail and to show how they contribute to give added value to the grey literature produced by ISFOL, by the SDC itself and by other public bodies in the framework of the European social fund for research in the field of vocational training and labour policies.

One of the services provided is the creation and implementation of the Catalogue of ISFOL documentation (which includes grey and published materials). It is an electronic database of all documentation produced by different ISFOL sectors since 1974 and collected in the multimedia library. It includes monographs, periodicals, and grey literature, in printed and electronic versions. Its bibliographic records include a description of the document following the ISBDs standards (including: Title and statement of responsibility; Edition, material or type of publication specification; publication, distribution, etc.; Physical description; Series; Notes, Standard number and conditions of availability) and its indexing (with the CDS thesaurus) and abstracting.

The catalogue is available online in the EUROPALAVORO web pages and can be searched for by date, title and keyword. The CDS has also made a CD-ROM version of the catalogue and every year it produces a printed version including the documents published in the last two years.

The paper version was originally intended to serve the needs of the library users but today has developed into a tool for the control of the ISFOL body of documentation.

In addition to the ISFOL documentation, in its multimedia library the CDS collects a wide range of documents on guidance, vocational training and labour policies. They include grey literature and published materials with special attention on documentation produced by organisations, at a national and European level, working in the same field of interest as ISFOL. The SDC also produces and makes available an online catalogue of this documentation (still on the EUROPALAVORO web pages). The LOGOS database includes bibliographic records of documents produced since 1998 and collected by the CDS. Besides the description of monographs, grey literature, periodicals and CD-ROM it also includes citations of articles from more than 140 titles of periodicals available in the multimedia library, all accompanied by an abstract.

The online database allows search by author (personal or corporate), by year, by title (or title word), by kind of material, by title of periodical and by keyword (selected from the CDS Thesaurus).

The opportunity of searching by kind of material is very important considering the characteristic of mixed catalogue of LOGOS. As a matter of fact it allows the selection of different kinds of GL materials.

From the CDS webpages it is also possible to search simultaneously on the ISFOL documentation and LOGOS databases.

The two databases together are an important tool to support social and economic research and the design and management of training activities.

While the two databases include both published and grey documents the SDC is also involved in activities specifically related to the promotion of grey literature control and availability.

Specifically the CDS participate for ISFOL to the SIGLE system (System for Information on Grey Literature in Europe).

The SDC, aware of the importance of grey materials' bibliographic control and availability for research, since 1998 has co-ordinated the ISFOL participation to the SIGLE through the CNR central library which is a national centre for Italy.

The SDC had an important role in spreading within the institute the importance of GL control at a European level and worked for the involvement of the different ISFOL areas in the project. As there are basic features grey materials must have to be included in the database, the participation to the SIGLE has also contributed to the standardisation of the ISFOL GL products. Twice a year the SDC transfers ISFOL GL bibliographic records to the CNR library. These records contain details of the cited document, (including authors names and the issuing organisation), document type, indexing and an abstract in English.

The SDC is also responsible for collecting grey documents (both in printed and electronic format) described in it and providing them on request of SIGLE users.

The CDS operate not only for organising, managing and distributing ISFOL documentation but is a GL producer itself. It co-ordinates the editing and publishing of monographs, serial materials, and multimedia products both at institutional level and in the framework its activities for European social fund.

At an institutional level, besides the already cited printed version of the catalogue of the ISFOL documentation, the CDS coordinate the editing of the INO ISFOL; an institutional monthly newsletter aimed at providing information on the activities of the institute. The full text of the newsletter is available online on the ISFOL website. It is also possible to register online to receive it monthly by email.

The CDS also produces and distributes the bimonthly review of European documentation and topical research EUROPA.DOC and EUROPA.DOC Speciale. Europa.doc is a bibliographic bulletin on European Union documentation. It aims to provide updated bibliographic information on research related to vocational training, guidance and labour policies. It focuses in particular on grey documentation produced by European Union institutions and agencies and by research organisations at a European level operating in the field. All documents cited in the bulletin are included in the LOGOS database and made available through the CDS specialised library.

Europa.doc speciale is a bibliographic bulletin of documentation on the Social dialogue in the “Mezzogiorno” (Area of South of Italy included in the Objective 1 of European Union) produced by the CDS in collaboration with the Fondazione Giulio Pastore in Rome. It includes updated bibliographic citations on social dialogue in the south of Italy especially on topics related to vocational training and labour policies. Also in this case special attention is given to grey documentation produced by EU institutions and other public and private organisations at a national and European level. Both Europa.doc and Europa.doc speciale are available in a printed version and in full text in the ISFOL website.

Within the project “Informazione e pubblicità per il FSE” (Information and publicity for ESF) the CDS produces also FSE news a monthly newsletter both in printed and electronic version available online in the Eurolavoro website. The newsletter is intended to inform on topics related to the ESF programming period 2000-2006. Besides the newsletter the Information and publicity project has implemented the FSE mail service. It provides updated information on ESF via email. It is an example of innovative means of distribution of information and documentation that reminds, in certain ways, about the traditional informal channels of communications used to distribute GL. It can be identified as a “new form of GL”.

The SDC is also responsible for the production of guides, monographs and multimedia products all of which are published outside commercial channels. Of particular interest, because it is also a work tool for the CDS, is the monograph “I termini della formazione: il controllo terminologico come strumento per la ricerca” published in the series “I libri del Fondo sociale europeo”. It contains the CDS thesaurus, a controlled indexing language formed by a list of terms on vocational training, guidance and labour. The volume is available online in full text on the Eurolavoro website.

Documentation produced by the CDS is intended as an instrument to improve information on the activities of the ISFOL and on its documentary production and at a wider level on documentation produced at institutional and European level on vocational training and labour policies.

We have seen how an important part of the activities carried on by the CDS is related to the dissemination of information on ISFOL and ESF activities and the distribution of its products.

Today the Internet is certainly the fastest and most efficient means for distributing information and documentation and the SDC fully exploit it as a main resource. But at the same time the importance that traditional printed documents have for many people as information source is recognised. For this reason electronic documents have not replaced printed materials.

## Conclusions

As the main public organisation operating in the field of vocational training and labour policies, ISFOL is a point of reference in Italy for researchers and training operators searching for information on these themes. At the same time because of its activities in support of the Ministry of Labour and Social Policies, Isfol has an important role in the dissemination of information on the European social fund in Italy.

“Public information sector”, information and documentation produced by public administrations has been recognised by the European Union as a key resource for Europe<sup>1</sup>. Public information in fact, intended as all documentation and information produced by public administrations, is a key element in many EU programmes and projects.

And grey literature is certainly the ideal means of expression for documentation produced by a public institute because of its lack of ties with commercial publishers and its free availability. This is even more true today that problems related to lack of visibility and difficulty of access of grey literature have been overcome by new means provided by information and communication technologies.

Today public organisations have at the same time the opportunity and the responsibility of providing wide access to the information and the documentation they produce.

ISFOL is aware of these new opportunities but also of the responsibility they carry with them.

And the SDC project is the result of this recognition. GL services provided by the SDC documentation produced within the activities of the institute more visible and widely available to researchers and training system operators.

---

## Author Information

### Isabella Pitoni

ISFOL

Visite di studio per specialisti della formazione

VIA MORGAGNI 33

00161 Rome

Italy

[i.pitoni@isfol.it](mailto:i.pitoni@isfol.it)

[d.macri@isfol.it](mailto:d.macri@isfol.it)

[www.isfol.it](http://www.isfol.it)

---

<sup>1</sup> Green paper on public sector information in the information society, COM (98)585 final, adopted on 20 January 1999 COM (2002) 207 definitivo - 2002/123 (COD)

## Management of Grey Literature for the construction of a database for the University of Maranhão's - UFMA and the State University of Maranhão - UEMA

Maria do Rosário Guimarães Almeida  
Cláudia Maria Pinho de Abreu Pecegueiro  
Georgete Lopes Freitas  
Raimunda de Jesus Ribeiro

**Abstract:** Recuperation of Grey Literature from the Federal University of Maranhão (UFMA) and at the State University of Maranhão (UEMA). Grey Literature contains information, which is of great importance to a considerable number of users. The volume of scientific material currently covered by this type of literature escapes the editorial circuits, and represents an extraordinary source of richness in unpublished information. The research is presented in three phases: the first phase, which is classified as descriptive, is the assessment of theses and dissertations of professors and technicians of the UFMA and UEMA universities. The second phase is the analysis and typing the variables: institution, author, title, instructor, library, area, resume and key words, for the purpose of feeding the database. The third phase aspires the electronic digitalization of the Digital Library Project in universities: a service offering the availability and accessibility of Grey documents (thesis/dissertations), indexed or to be indexed at UFMA and at UEMA which enables the identification of the producers; establishes norms and technical standards for the control, the divulgence and transfer of Grey Literature produced at UFMA and which can be accessed through the System of Grey Literature in Europe (SIGLE), Information Center for Latin America and the Caribbean (BIREME), the Brazilian Institute of Scientific Information and Technology (IBICT) and the *Networked Digital Library of Thesis and Dissertations* (NDLTD). Ninety-three typed documents were found at UFMA concerning: Human Science, Health Science, Engineering, Exact Science, Biological Science, Applied Social Agrarian Science, Linguistics, Language Arts, and Art, corresponding to 27,27% of the total number of Grey documents verified from the listing that was made available by the Technical Processing of UFMA. The largest volume of Grey Literature typed was in the area of Electrical Engineering with 9,92%, in relation to UEMA which had forty-three documents typed out, this number corresponds to 35% of the total number of Grey documents found at UEMA. Data shows that the largest register of documents found is in the area of Veterinary Science. All documents registered were also found in paper form.

**Key Words:** Grey Literature; Grey Literature - Research Project; Grey Literature - Federal University of Maranhão; Grey Literature - State University of Maranhão.

### 1 JUSTIFICATION

As a starting point, the scientific knowledge generates trust by the techniques and methods used, which divulge the research in progress and guarantee the continued production of grey literature.

The formal channels of scientific communication cover publications of greater divulgation, such as books and periodicals; the informal channels are merely informative of personal nature or which refer to current research, such as congress assignments and of similar characteristics.

Thus, all research involves diverse activities of communication and at least produces formal communication. Actually, certain research topics generally produce several publications, which are generated during the accomplishment finalization of such. These publications vary in **format** [reports, congress assignments, speeches, periodical articles, books, among others], **audiences**

[colleges, students, general public] and **function** [inform, obtain reactions, register authorship, indicate and locate documents, etc.] (ZIMAN apud MUELLER, 2000, p. 22, author’s griffon).

There’s a divergence among the scientific community as to the classification of the formal, informal, semi-informal and intermediate channels of communication. Among the semiformal channels of dissemination of knowledge, the theses and dissertations, in particular are the inserts of the one of the types of Grey Literature, as Almeida conceptualizes (2002, p.37).

[...] the conjunction of documents, independent of their typology and support, or printed or electronic format, emitted by university centers, research centers, companies, industries, public and private academic societies, and of no intention of being published. These are of vital importance in the transfer of knowledge.

The highlight of this study is building a multidisciplinary database in the scientific areas of knowledge seeking to collect the recuperated UFMA and UEMA Grey Literature (initially theses and dissertations), from 1998 to 2003 collection. This data is also available for access through the System of Grey Literature in Europe (SIGLE), Information Center for Latin America and the Caribbean (BIREME), the Brazilian Institute of Scientific Information and Technology (IBICT) and the *Networked Digital Library of Thesis and Dissertations* (NDLTD).

This report is based on descriptive research, starting from the theoretic fundaments to data analysis.

## 2. GREY LITERATURE

Grey Literature is known by this terminology since 1980, “[...] it can be compared with the Grey mass inside the brain, characterized by its intellectuality”. (WILLMOTT, 2002, p.1). Due to such particularity, it is considered a document of utmost value to a considerable number of users, because it constitutes prime material of specialized databases.

Such literature should be classified by document categories, such as, congress minutes, norms, reports, patents, commercial literature, thesis, bulletins, official publications, etc. Grey Literature contrasts white documents in the following areas: Grey Literature is not available through formal or commercial channels, which makes difficult the access to such limited publications, and causes delay in converting it into books or periodicals.

The growth and usage of such information technology in the scientific investigations and also in the electronic publications allowed the arrival of new genres of documents, some being produced at the margin of the commercial circuits. Specifically speaking of Grey Literature, the data and resources available through the Internet are available through *FTP* or *WAIS*, *Newsgroups*, *Discussion Lists* stand out, Luzi (apud CORREA 2000), as well as preliminary versions of articles of periodicals and book chapters, communication of scientific conferences, *software*, databases and electronic compilations.

The technological development on the one hand also brought new opportunity in the field of administrative organization, divulgence and storage of Grey documents. Diverse projects and incentives are found randomly with the objective of promoting this type of publication, some of which are relevant such as *Greynet*, which prioritizes the access to reports of scientific investigation and promotes the international congress, (the next being *International Conference on Grey Literature*, 5, of 4-5 of December 2003, in Amsterdam) and a *Networked Digital Library of Thesis and Dissertations* (NDLTD), which deals with a digital library of thesis and dissertations, *Scholarship* University of California these are the digital storage places especially directed to the publication of preliminary versions of scientific documents elaborated by the staff connected to the promoting academic institutions .

Grey Literature, by its unique form of divulgence has the advantage of supplying updated information to the scientific community in a rapid manner, although it is at times only available to the surrounding community and a limited number of users.



### 3 OBJECTIVES

The objectives of research are constituted by the general and specific categories.

#### 3.1 General

Build a database of thesis and dissertations from UFMA and UEMA to feed international databases as the System of Grey Literature in Europe (SIGLE) and the Center of Information for Latin America and the Caribbean (BIREME), the Brazilian Institute of Scientific Information and Technology (IBICT) and the *Networked Digital Library of Thesis and Dissertations* (NDLTD).

#### 3.2 Specific

- a) identify the units that produce Grey Literature at UFMA and UEMA;
- b) establish the norms and standards of bibliographical description and indexation for Grey Literature, compatible with other existing systems;
- c) create and implant norms for the recuperation, analysis, dissemination, access and usage of Gray Literature produced at UFMA and UEMA;
- d) build a program of database elaboration (dissertations and thesis);
- e) develop mechanisms which enable the interface of communication with other similar units of production, treatment and divulgence of Grey Literature.

### 4 METHODS AND PROCEDURES

The research is divided into three phases: The research is presented in three phases: the first phase, which is classified as descriptive, is the assessment of theses and dissertations of professors and technicians of the UFMA and UEMA universities. The second phase is the analysis and typing the variables: institution, author, title, instructor, library, area, resume and key words, for the purpose of feeding the database. The third phase aspires the electronic digitalization of the Digital Library Project in universities: a service offering the availability and accessibility of Grey documents (thesis and dissertations), indexed or to be indexed at UFMA and at UEMA which can be accessed by the System of Grey Literature in Europe - SIGLE and the Center of Information for Latin America and the Caribbean - BIREME and the Brazilian Institute of Scientific Information and Technology (IBICT) and the *Networked Digital Library of Thesis and Dissertations* (NDLTD).

#### 4.1 The universe of research

The universe of research is constituted of different types of Grey Documents available at UFMA and UEMA.

#### 4.2 Material

The Grey Literature generated at UFMA and UEMA is limited to thesis and dissertations.

#### 4.3 The gathering of data

For the accomplishment of this research paper, the following didactic procedures were used:

- a) research of literature to access available documentation;
- b) field research using the following steps:
  - First, Researchers, Course Coordinators, Head of Departments and units, or post-graduation nucleuses at UFMA and UEMA; were contacted, in order to locate and recover the Grey Literature available. Next, the format by which the Grey Literature was presented was identified. Then, from a meeting was held with the researchers and technicians involved in the research aiming to build a program for the elaboration of databases. And finally, quantifying the Grey Literature through bibliometric studies.
- c) After determining the objectives of the research, the data collected from thesis and dissertations from UFMA and UEMA related to the period of 1998- 2003 were analyzed.

## 5 RESULTS AND DISCUSSION

What stands out in the first phase is that the data collected corresponds to the documents pertaining to UFMA and UEMA.

The partial results of the research were divided into three phases for the development of actions:

- a) bibliographical research about the current subject for greater profounding in the theoretic reference about Grey Literature;
- b) verification of the dissertation listings and thesis through a listing obtained from the Central Library Report of UFMA;
- c) the third phase is the updating the database of dissertations found at the Central Library, in the Technological Sector and in the Post-graduation departments of Social Science at the Federal University of Maranhão.

Although the database of the Post-graduation Department of UFMA does not contain updated information concerning personnel in the Grey Literature area, partial data was obtained through the Automation System of the Human Resources Department.

For this study it is emphasized that the High school University is categorized as a Department (Frame A).

Location of Sector	TITLE OF PROFESSORS AT UFMA			
	Specialization	Doctor's	Master's	Total /Department
University High school	47	00	06	53
Department of Arts	03	03	05	11
Department of Philosophy	07	00	15	22
Department of Language Arts	09	04	11	24
Department of Psychology	02	00	14	16
Department of History	03	00	14	17
Department of Geosciences	05	01	09	15
Department of Sociology and Anthropology	00	15	12	27
Department of Library Science	02	01	14	17
Department of Economy	01	04	14	19
Department of Social Communication	07	02	14	23
Department of Law	11	01	14	26
Department of Law in Imperatriz	03	00	00	03
Department of Education	04	01	10	15
Department of Education I in Imperatriz	01	00	01	02
Department of Education II	04	05	13	22
Department of Education II in Imperatriz	01	00	00	01
Department of Social Service	04	04	14	22
Department of Accountancy and administration	08	00	04	12
Department of Accountancy in Imperatriz	01	00	00	01
Department of Morphology	15	01	04	20
Department of Physical Education	05	05	02	12
Department of pharmacy	13	02	10	25
Department of Pharmacy	06	03	08	17
Department of Medicine	08	08	09	25
Department of Medicine II	20	03	03	26
Department of Medicine III	10	08	07	25
Department of Deontology	02	03	12	17
Department of Deontology II	01	03	15	19
Department of Pathology	06	03	11	30
Department of Public Health	06	04	11	21

Department of Biology	01	14	03	18
Department of Physiologic	05	03	13	21
Department of Oceanography and Limnology	02	11	02	15
Department of Chemistry	02	10	08	20
Department of Mathematics	04	04	10	18
Department of Design and Technology	04	03	07	14
Department of Technology and Chemistry	00	09	09	18
Department of Electrical Engineering	01	13	16	30
Department of Physics	06	14	03	24
Department of Computer Science	03	03	10	16
Pro-rector of Learning	01	00	00	01
Imperatriz Campus I	01	00	00	01
<b>Subtotal</b>	<b>245</b>	<b>178</b>	<b>357</b>	<b>780</b>
<b>Total</b>	<b>780</b>			

Frame A - Number of Teachers at UFMA

Source: FEDERAL UNIVERSITY OF MARANHÃO. Human Resources Department, 2003

From the total of 43 (forty-three) existing departments at UFMA, the University High school was the department which registered the highest number of specialists: 19,18%; the greatest number of Doctor's Degrees is in the Department of Sociology and Anthropology, representing 8,42% of the professors at this graduation level at UFMA; the Department of Electric Engineering had the highest number of masters, totaling 4,48% of the masters of UFMA in this Department. 25,48% of the Departments do not have Doctor Professors in their group of teachers, and 4,65% of the Departments do not have Specialist Professors, and 11,62% do not have masters.

After locating the Grey Literature, the data must be typed in a spreadsheet (Appendage A), built by research members, whose field of work is related to the areas of knowledge of the document described.

The classification by the National Counsel of Scientific and Technological Development - CNPq has been adopted, for the areas of knowledge and from the Brazilian Association of Technical Norms - ABNT and Registered Brazilian Norm - NBR 6030/2002, as reference.

From the period of October 2002 to April 2003 we concentrated on the UFMA documents. 83 (Eighty-three) documents in the areas of Exact Science and Technology were listed, 124 (one-hundred and twenty-four) in the area of Health, 134 (one-hundred and thirty-four) in the area of Social and Human Science, totaling 341 (three hundred and forty-one) recuperated documents (by a listing made in a report conceded by the Technical Processing of UFMA) pertinent to the research, distributed into (Frame B):

AREA	INDEXED
EXACT SCIENCE AND TECHNOLOGY	83
HEALTH	124
SOCIAL AND HUMAN SCIENCE	134
TOTAL	341

Frame B - Amount of thesis and dissertations

The amount of thesis and dissertations detailed in (Frame B) demonstrates the total per area, obtained through a report requested from the Technical Processing of the Central Library of UFMA. It demonstrates that the area of Social and Human Science; 39,29% possess a greater percentage of dissertations and thesis, and the area of Exact Science and Technology a lower number, 24,34%. The total of thesis and dissertations, do not correspond to the grand total, totaling 780 (seven hundred and eighty) professors. Until the present moment, 78 (seventy-eight) dissertations have been typed and 15 (fifteen) thesis (Frame C) pertaining to the areas of:

Human Science Health Science, Engineering, Exact Science, and Earth Science, Biological Science, Applied Social Agrarian Science, Linguistics, Language Arts and Art, Corresponding to 27,27% of the total number of Grey Documents covered in the listing made available by Technical Processing of UFMA (Frame D).

AREA	DISSERTATIONS	THESIS	%
Human	01	00	1,08
Health	24	11	37,63
Engineering	36	01	39,78
Exact and Earth	07	01	8,60
Biological	01	02	3,23
Agrarian	00	00	0,00
Applied Social	09	00	9,68
Linguistics, Language Arts and Art	00	00	0,00
<b>TOTAL</b>	<b>78</b>	<b>15</b>	<b>100</b>

Frame C - Percentage of typed documents

By analyzing Frame C, the greatest number registered is in Health Science, 39,29%. However, the category that contains the largest number of typed Grey Literature in Engineering with 9,92%. One hundred percent of the documents collected were also in paper form.

To store the Grey Literature according to the norms and bibliographical description, a interface of communication was developed in conjunction with the Department of Electrical Engineering for the divulgence of Grey Literature at a regional, national and international level.

## 6 CONCLUSION

Grey Literature is a rich source of unpublished information which is now being recovered by digital libraries at private and public universities.

Although university departments have joined together to gather such information, there was a certain difficulty in locating the necessary data for the construction of local and international databases. This is now being overcome by the fact that the more recent Grey Literature is being developed and published in digital form.

With the continued feeding of information and expansion of network users, the Grey Literature channel is considered the fastest and most updated source of knowledge in the scientific community.

## REFERENCES

ALMEIDA, Maria do Rosário Almeida. *Literatura cinzenta...* São Luis: UFMA/Souzandrade, 2000.

ÁLVAREZ-OSSORIO, José R. Pérez. *Introducción a la información y documentación científica*. Madrid: Alhambra, 1988. 107p.

ASSOCIATION FRANÇAISE DE NORMALIZATION. *Vocabulaire de la Documentation*. Paris, 1983. [Projet de norme enregistrée].

BRADFORD, Samuel Clement. *Documentação*. Rio de Janeiro: Fundo de Cultura, 1961. 292p.

BRIET, Suzanne. *¿Qué es la documentación?* Santa Fé: Universidad Nacional del Litoral, 1960. 48p.

CARROL, Bonnie C., COTTER, Gladys A. A new generation of grey literature: the impact of advanced information technologies. In: INTERNATIONAL CONFERENCE ON GREY LITERATURE, 1., 1993, Amsterdam, *Conference Proceedings...* Amsterdam: TransAtlantic, 1994. 393p.

CHARTIER, Roger. Do código ao monitor: a trajetória do escrito. *Estudos Avançados*, São Paulo, v. 8, n.21, maio/ago.1994.

COMBEROUSSE, Martine. Les nouvelles technologies au service de la littérature grise. *Bulletin des bibliothèques de France*, Paris, v.40, n.2. 1995.

CURRÁS, Emilia, HENRICH, Norbert. **Estructuración de un programa de sistema experto e hipertexto para reconocimiento y localización de ciertos documentos de literatura gris.** Madrid, [1993]. Pretirada.

DI CESARE, Rosa, SALA, Cesare. The use of grey literature in the agricultural economics field: a quantitative analysis. In: INTERNATIONAL CONFERENCE ON GREY LITERATURE, 2. , Washington, 1995. **Conference Proceedings...** Washington, D.C. Transatlantic/ GreyNet, 1995. 265p.

FARACE, Dominic. Grey Literature and publishing. **NewsBriefNews**, Amsterdam, TransAtlantic/ GreyNet, v.7, n.1. 1998

FONSECA, Edson Nery da. **Introdução à Biblioteconomia.** São Paulo: Pioneira, 1992. 286p.

GIBB, J. M., PHILIPPE, E. A better fate for the grey or non-conventional literature. **Journal of research communication studies**, v.1, p.225-234.

MALCLÈS, Louise-Noëlle. **Les sources du travail bibliographique.** Genève: Lille Libraire, 1950. v.1.

MARTÍN VEGA, Arturo. **Fuentes de información general.** Gijón: Trea, 1995. 320 p.

PEREIRA, Maria de Nazaré Freitas. **Tecnologia da informação em redes eletrônicas de computadores.** Rio de Janeiro, 1995. 178p. [Apostila]

KATZ, William A. **Introduction to Reference Work.** New York: McGraw-Hill, 1972. v.2.

UNION FRANÇAISE DES ORGANISMES DE DOCUMENTATION. **Cours techniques de documentation.** Paris: Documentatio générale, 1951.

VILLASEÑOR, Isabel. **Las fuentes de información I y II** .In: LOPEZ YEPEZ, José. (Coord.). **Manual de Documentación y Información.** Madrid: Piramide, 1996. 541p.

ZAHER, Célia Ribeiro. **Introdução à documentação.** 2.ed. Rio de Janeiro, 1988. 184p.

---

## Author Information

### **Dr. Maria do Rosário Guimarães Almeida**

Phd in Information Science – Project Coordinator - FAPEM Researcher.

Universidade Federal do Maranhão - UFMA

Rua dos Beijas-Flores, Q 15, casa 24, Ponta do Farol

65077-150 Sao Luis - Maranhão

Brazil

mrga@elo.com.br

## Appendices

List of Participating Organisations	42-43
Publication Order Form	44
Index of Authors	45

## List of Participating Organisations

British Library - BL	United Kingdom
BZ&S Accountants and Advisors Ltd.	Netherlands
Centre for Information on Low External Input and Sustainable Agriculture - ILEIA	Netherlands
Centre National de Recherche Scientifique - CNRS	France
City of Amsterdam, Bureau External Affairs	Netherlands
Columbia University Press	United States
Consiglio Nazionale delle Ricerche - CNR	Italy
Dalhousie University	Canada
Emporia State University	United States
Environmental Conservation Branch, Environment Canada Atlantic Region	Canada
European Association for Grey Literature Exploitation - EAGLE	Germany
European Centre for Higher Energy Physics - CERN	Switzerland
Federal University of Maranhão - UFMA	Brazil
Grey Literature Network Service - GreyNet	Netherlands
Helsinki University of Technology - HUT	Finland
Horizons	United States
Information Processing Centre - OPI	Poland
Informationszentrum Sozialwissenschaften - IZ	Germany
Institut de l'Information Scientifique et Technique - INIST	France
Institute of Medical Psychology and Education - IMP/E	France
Istituto per lo Sviluppo della Formazione dei Lavoratori - ISFOL-CDS	Italy
Istituto Superiore di Sanità - ISS	Italy
Istituto Superiore per la Prevenzione e la Sicurezza del Lavoro - ISPESL	Italy
Japan Science and Technology Agency - JST	Japan
JVL Industri Elektronik	Denmark
Koninklijke Bibliotheek - KB	Netherlands
Koninklijke Nederlandse Akademie van Wetenschappen - KNAW	Netherlands

---

Latvian Academic Library - LAL	Latvia
Library of Zeeland	Netherlands
LINK s.r.l.	Italy
National Diet Library - NDL	Japan
Nederlands Instituut voor Wetenschappelijke Informatiediensten - NIWI	Netherlands
New York Academy of Medicine - NYAM	United States
Office of Scientific & Technical Information, Department of Energy - OSTI/DOE	United States
Organon NV	Netherlands
Portland State University, Branford P. Millar Library	United States
Scientific and Technical Information Centre of Russia - VNTIC	Russia
Tilburg University Library - UvT	Netherlands
United Nations Educational, Scientific and Cultural Organization - UNESCO	France
Université des sciences et technologies de Lille, Bibliothèque centrale	France
University of Amsterdam - UvA	Netherlands
University of Bergen, Research Documentation Unit - UIB	Norway
University of California, Irvine - UCI	United States
University of Pretoria - UP	South Africa
University of Twente - UTwente	Netherlands
Utrecht University Library - UU	Netherlands
Wageningen University Library - WUR	Netherlands



## Index to Authors

Almeida, M.	152	MacDonald, B.H.	50
Artus, H.M.	1	Macri, D.	124, 147
		Manco, M.	40
Baron, T.	106	Minderhoud, M.	69
Blaaij, C. de	112		
		O'Dell, C.	117
Castriotta, M.	40		
Claerebout, M.	27	Pavlov, L.P.	64
Cordes, R.E.	50	Pitoni, I.	124, 147
Dallman, D.	117	Roem, W.	69
De Castro, P.	32	Roosendaal, H.E.	75
Di Cesare, R.	40		
		Salinetti, S.	32
Farace, D.J.	10	Schöpfel, J.	94
Frantzen, J.	10	Siegel, G.E.	13
		Simko, T.	106
Gelfand, J.	21	Stock, C.	94
Hofe, H. von	132	Théaux, W.	140
Kairamo, A.	87	Vesely, M.	106, 117
		Vigen, J.	117
Le Meur, J.	106		
Libutti, L.	40	Wells, P.G.	50
Luzi, D.	40		