

# New Frontiers in Grey Literature

## Sponsors

**BIOSIS** Publisher of Biological Abstracts and Zoological Record

**JST** Japan Science and Technology Corporation

**MCB** University Press  
Networked Publisher in  
Management, Engineering and  
Library Science

**NASA** National Aeronautics  
and Space Administration

**NLE** National Library of  
Education, United States  
Department of Education

## Program Committee

**BIOSIS** Publisher of Biological Abstracts and Zoological Record

**EAGLE** European Association  
for Grey Literature Exploitation

**FID** International Federation for  
Information and Documentation

**GreyNet**  
Grey Literature Network Service

**IFLA** International Federation  
of Library Associations and  
Institutions

**JST** Japan Science and  
Technology Corporation

**MCB** University Press  
Networked Publisher in  
Management, Engineering and  
Library Science

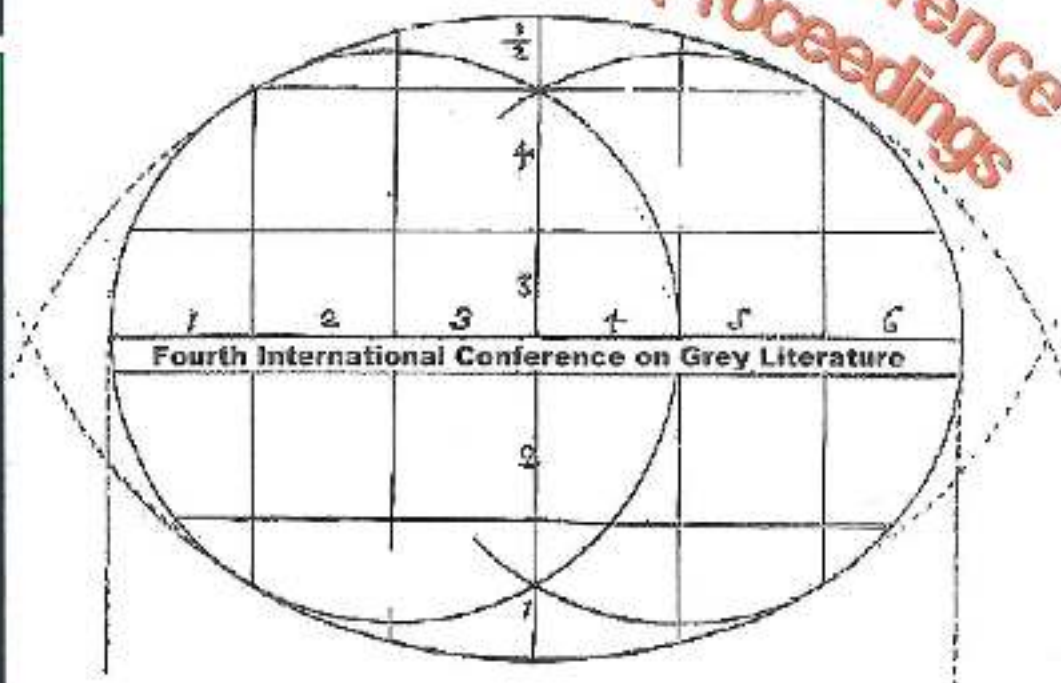
**NASA** National Aeronautics  
and Space Administration

**NLE** National Library of  
Education, United States  
Department of Education

# GL'99

Kellogg Conference Center  
Washington D.C., USA  
4-5 October 1999

Conference  
Proceedings



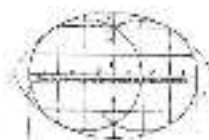
## Program and Conference Bureau:

*GreyNet*, Grey Literature Network Service  
Koninginneweg 201, 1075 CR Amsterdam, The Netherlands  
Tel/ Fax: 31-20-671.1818 - [GreyNet@inter.nl.net](mailto:GreyNet@inter.nl.net)  
<http://www.konbib.nl/infolev/greyNet>





**New Frontiers in Grey Literature**  
**FOURTH INTERNATIONAL CONFERENCE ON GREY LITERATURE**



● Reports ● Proceedings ● Dissertations ● Bibliographies ● Manuals  
● CD-ROM ● Databases ● Bulletin Boards ● WWW ● Email ● Internet ● Etc.

Kellogg Conference Center

Washington D.C., USA

4-5 October 1999



**GL'99 Program and Conference Bureau**

GreyNet, Grey Literature Network Service

Koninginneweg 201, 1075 CR Amsterdam

The Netherlands

Phone/Fax : 31-20-671.1818

Email : GreyNet@inter.nl.net

<http://www.konbib.nl/infolev/greynet/GL'99.htm>

**Conference Proceedings on Grey Literature : GL-conference series /**  
compiled by D.J. Farace ; with the assistance of J. Frantzen. - Amsterdam :  
GreyNet, Grey Literature Network Service. - ISSN 1386-2316

No. 1: GL'93 Conference Proceedings - 416 p. - 1994. - ISBN 90-74854-03-6 : € 55

No. 2: GL'95 Conference Proceedings - 266 p. - 1996. - ISBN 90-74854-08-7 : € 55

No. 3: GL'97 Conference Proceedings - 294 p. - 1998. - ISBN 90-74854-17-6 : € 55

No. 4: GL'99 Conference Proceedings - 322 p. - 2000. - ISBN 90-74854-24-9 : € 55

## CIP

### **GL'99 Conference Proceedings**

Fourth International Conference on Grey Literature : GL'99 Conference Proceedings :  
New Frontiers in Grey Literature, Washington D.C., 4-5 October 1999 / compiled  
by D.J. Farace ; with the assistance of J. Frantzen. - Amsterdam : GreyNet,  
Grey Literature Network Service, January 2000. - XII, 322 p. ; 30 cm. - Index. -  
(GL-Conference series, ISSN 1386-2316 ; No. 4)

*Contains the full text of the twenty-eight papers presented during the Opening, Plenary and  
Breakout Sessions*

ISBN: 90-74854-24-9

© 2000, GreyNet



## Acknowledgements

On behalf of the Sponsors and Program Committee Delegates, it is my distinct pleasure to welcome you, the authors and participants to this, the Fourth International Conference on Grey Literature. You have travelled here, to the capital of the United States of America, from more than twenty countries worldwide. Together, you represent more than one hundred organisations in the various sectors and levels of government, academics, business and industry. We have been brought together by the common theme "New Frontiers in Grey Literature", to discuss new worlds of topics, formats, uses and applications of Grey Literature - both in print and electronic formats, not controlled by commercial publishers. Nevertheless, we hold this information to be valuable - responsible often in the first place for that which later only becomes recognized in small measure. Our frontiers cross time and space. Temporal, because we are on the eve of a new century, a new millennium. Spatial, belonging to geographical, regional, as well as, virtual space and locators.

We are here not as one, but as many and varied; however, we share a common vision, a noble mission - not as pioneers but instead as that which we are:

- Librarian and Documentalists,
- Professors and Lecturers,
- Editors and Publishers,
- Managers and Policy Makers,
- Project Coordinators & Database Administrators,
- Journalists and Business Analysts,
- Researchers and Information Specialists,
- Etcetera.

No, we are not one, however we are truly united, absolutely diversified. We here, span fields and cross the disciplines of agriculture, rural development, aerospace, applied sciences, military and defence, physics, social sciences, economic, computer science, transportation studies, medicine, education, management and engineering, political science and international affairs, biology, zoology, energy, environment, and last but not least, library and information science.

Yes, you are all truly welcome to GL'99. Your one and real mission - not impossible, but instead inspired - to demonstrate at this international forum, the results of 49 authors and co-authors months of research and writing. Bridging theoretical constructs and academic terminology with empirical facts, models, and paradigms.

At this fourth international conference, perhaps a bit more than at the past three in this series, we stand at the crossroads, not on the periphery, but at the very heart of this grey matter. Now, ready more than ever to demonstrate that grey literature is not merely the document in electronic and print format, but is that which encompasses the very process of communication responsible for knowledge transfer - encoded and deciphered.

**Dr. Dominic J. Farace**  
Program and Conference Director

## Conference Sponsors



**BIOSIS**  
Publisher of Biological Abstracts and Zoological Record



**JST**  
Japan Science and Technology Corporation



**MCB University Press**  
Networked Publisher in Management, Engineering, and  
Library Science



**NASA**  
National Aeronautics and Space Administration



**NLE**  
National Library of Education,  
United States Department of Education

## Program & Conference Bureau

*GreyNet*

Grey Literature Network Service  
Koninginneweg 201  
1075 CR Amsterdam, The Netherlands

Tel/Fax: 31-671.1818  
Email: [GreyNet@inter.nl.net](mailto:GreyNet@inter.nl.net)  
URL: <http://www.konbib.nl/infolev/greynet/GL'99.htm>

## Program Committee

<b>BIOSIS</b>	<b>Bruce H. Kiesel, Director IAD</b> Publisher of Biological Abstracts and Zoological Record
<b>EAGLE</b>	<b>Andrew Smith, Chairman Technical Committee</b> European Association for Grey Literature Exploitation
<b>FID</b>	<b>Martha B. Stone, President</b> International Federation for Information and Documentation
<b>GREYNET</b>	<b>Dominic J. Farace, Head</b> Grey Literature Network Service
<b>IFLA</b>	<b>Graham P. Cornish, Director UAP</b> International Federation of Library Associations and Institutions
<b>JST</b>	<b>Masayuki Sato, Director Washington Office</b> Japan Science and Technology Corporation
<b>MCB</b>	<b>Eileen Breen, Managing Editor</b> MCB University Press
<b>NASA</b>	<b>Carolyn E. Floyd, Information System and Services Division</b> National Aeronautics and Space Administration
<b>NLE</b>	<b>E. Stephen Hunt, Director for Planning and Policy</b> National Library of Education; U.S. Department of Education

## Conference Program

### Opening Session

Chairperson: Dominic J. Farace, GreyNet  
Grey Literature Network Service

### Session One

Global Assessment of Grey Literature  
*A Brave New World of Topics, Formats, and Uses*

Chairperson: E. Stephen Hunt, NLE  
National Library of Education, United States Department of education

### Session Two

Publishing and Archiving Electronic Grey Literature  
*From Production to Full-Text Storage, Retrieval and Distribution*

Chairperson: Eileen Breen, MCB University Press  
Networked Publisher in Management, Engineering and Library Science

### Session Three

Copyright and Grey Literature  
*Authorship, Ownership, and Property Rights*

Chairperson: Julia Gelfand, UCI  
University of California, Irvine

### Breakout Sessions:

#### Track A:

Policy and Management of Grey Literature  
*Chairpersons: Bruce H. Kiesel, BIOSIS and Masayuki Sato, JST*

#### Track B:

Applications and Uses of Grey Literature  
*Chairpersons: Carolyn E. Floyd, NASA and Andrew Smith, EAGLE*

### Product and Service Reviews

### Closing Session

Chairperson: Keith G. Jeffery, CLRC  
Central Laboratory of the Research Councils



# Table of Contents

## Acknowledgements

iii

## Opening Session

Chair: Dominic J. Farace, GreyNet, Grey Literature Network Service

Julia M. Gelfand - United States ..... 1-11

University of California, Irvine

*Grey Literature : Plausible Applications for Distance Education Practices*

Helmut M. Artus - Germany ..... 12-24

Social Science Information Centre, IZ

*The Internet and the Socio-Structural Change of Informal Scientific Communication*

L. Olatokunbo Aina - Botswana ..... 25-31

University of Botswana, UB

*Grey Literature and Library and Information Studies (LIS) : A Global Perspective*

## Plenary Session One

### Global Assessment of Grey Literature:

A Brave New World of Topics, Formats, and Uses

Chair: E. Stephen Hunt, NLE, National Library of Education, U.S. Department of Education

Gail McMillan - United States ..... 32-46

Virginia Polytechnic Institute and State University

*Perspectives on Electronic Theses and Dissertations*

Bonnie A. Osif - United States ..... 47-60

Pennsylvania State University

*International Transportation Literature : An Analysis of Citation Patterns, Availability and Research Implications to the Transportation Community*

Angela Aceti, Maria Castriotta,

Rosa Di Cesare, and Daniela Luzi - Italy ..... 61-77

ISPESL, Istituto Superiore per la Prevenzione e la Sicurezza del Lavoro;

CNR, Consiglio Nazionale delle Ricerche

*A profile of GL producers in the field of safety and health at workplaces in Italy : results of a sample survey*

## Plenary Session Two

Publishing and Archiving Electronic Grey Literature:  
From Production to Full-Text Storage, Retrieval and Distribution

Chair: Eileen Breen, MCB University Press

Michael L. Nelson and Kurt Maly - United States ..... 78-87 10 P  
 NASA Langley Research Center, LaRC ; Old Dominion University, ODU  
*Preserving the Pyramid of STI Using Buckets*

Keith G. Jeffery - United Kingdom ..... 88-108 21 P  
 Central Laboratory of the Research Councils  
*An Architecture for Grey Literature in a R&D Context*

Deborah E. Cutler - United States ..... 109-115 7 P  
 Office of Scientific & Technical Information, OSTI  
*Grey Literature in Energy : A Shifting Paradigm*

Elena Lodi - Italy ..... 116-134 19 P  
 Martin Vesely and Jens Vigen - Switzerland  
 University of Siena ; European Centre for High Energy Physics, CERN  
*Link Managers for Grey Literature*

## Plenary Session Three

Copyright and Grey Literature:  
Authorship, Ownership, and Property Rights

Chair: Julia Gelfand, University of California, Irvine

Michael Seadle - United States ..... 135-142 8 P  
 Michigan State University, MSU  
*Grey Copyrights for Grey Literature : National Assumptions, International Rights*

Cees A.T. de Blaaij - The Netherlands ..... 143-155 13 P  
 University of Amsterdam, UvA - Institute for Information Law, IVIR  
*Intellectual property on the move : Some observations on authors' rights, grey literature, publishers and last but not least access to information in a cyberian age*

Dave Davis - United States ..... 156-163 8 P  
 Copyright Clearance Center  
*The Application of Electronic Copyright Management Systems (ECMS) to Grey Literature*

Graham P. Cornish - United Kingdom ..... 164-169 6 P  
 International Federation of Library Associations and Documentation, IFLA  
*Copyright : Black and White or just making you see Red?*



Breakout Session - Track A	Policy and Management of Grey Literature
----------------------------	--

## Chairpersons:

Bruce H. Kiesel, Publisher of Biological Abstracts and Zoological Records, BIOSIS  
 Masayuki Sato, Japan Science and Technology Corporation, JST

Ichiko T. Morita - United States ..... 170-180 11

Japan Documentation Center ; Library of Congress

*Issues in Distribution of Grey Literature : Experience of the Japan  
 Documentation Center (JDC)*

Maria do Rosário Guimarães Almeida - Brazil ..... 181-198 18p

Univesidade Federal do Maranhão, UFMA

**In Absentia:** *Control and Access for Grey Literature in Brazil : A Proposal*

Joachim Schopf - France ..... 199-202 4

Institut de l'Information Scientifique et Technique, INIST-CNRS

*INIST : Tracking Grey Literature in a Changing Environment*

Atefeh S. McCampbell,

Linda M. Clare, and Scott H. Gitters - United States ..... 203-216 14

Florida Institute of Technology ; Medical System Information Technology Group ;

Lippincott Williams & Wilkins

*The Competitive Advantage for the New Millennium : Knowledge Management*

Hyun-Yang Cho, Seon-Heui Choi,

Sung-Pil Choi, and Huyng-Sup Shim - Korea ..... 217-226 10

Korea Research and Development Information Center, KORDIC

*Archiving Electronic Grey Literature of Korean Academic Societies  
 in Science and Technology*

Breakout Session - Track B	Applications and Uses of Grey Literature
----------------------------	--

## Chairpersons:

Carolyn E. Floyd - National Aeronautics and Space Administration, NASA

Andrew Smith - European Association for Grey Literature Exploitation, EAGLE

Linda R. Musser - United States ..... 227-229 3p

Pennsylvania State University, PSU

*Mine Maps as Grey Literature*

<b>Breakout Session - Track B</b>	<b>Applications and Uses of Grey Literature (continued)</b>
-----------------------------------	---

Jean-Yves Le Meur and David Dallman - Switzerland	230-237	89
European Centre for High Energy Physics, CERN		
<i>Automatic Keywording of High Energy Physics</i>		
Fethy Mili - Canada	238-253	169
University of Montreal		
<i>Trends in Publishing Academic Grey Literature : Examples form Economics</i>		
Krystyna Siwek - Poland	254-259	69
Information Processing Centre, OPI		
<i>Grey Literature Based Surveys of Information Industry Development in Central and Eastern Europe</i>		
T. Rama Devi and Anil Takalkar - India	260-275	169
Centre on Rural Development, National Institute of Rural Development		
<i>Use Pattern of Grey Literature in Rural Development : An Indian case study</i>		
Diane Helmer, Isabelle Savoie and Carolyn Green - Canada	276-281	69
British Columbia Office of Health Technology Assessment, BCOHTA		
<i>How do various Fugitive Literature Searching Methods Impact the Comprehensiveness of Literature Uncovered for Systematic Review?</i>		
Hiroyuki Sato, Kenshou Nakagawa and Yoshiyuki Maeda - Japan	282-296	159
Japan Science and Technology Corporation, JST		
<i>Directory Database of Research and Development Activities (ReaD)</i>		
Rosa Di Cesare and Giovanni Lazzari - Italy	297-306	109
Consiglio Nazionale delle Ricerche, CNR ; Biblioteca della Camera dei Deputati		
<i>Towards integration of information sources on grey literature : A Case Study</i>		
Halima Semra - Algeria	307-313	79
University of Constantine		
<b>In Absentia:</b> <i>Internet use : Research of Grey Literature by teachers-researchers at the University of Constantine</i>		

## Appendices

List of Participating Organizations	315
Author Index	319
GreyNet Membership 2000	320
Publication Order Form	322



# GreyNet

Grey Literature Network Service

*Your Gateway to Grey Literature*

**Newsletters**

**Bibliography**

**Conferences**

**Welcome**

**R & D**

**Publications**

**Guide**

GL'99

What's New

*GreyNet* a branch of **MCB**  
University Press

**GREY LITERATURE NETWORK SERVICE**  
Koninginneweg 201, 1075 CR Amsterdam, Netherlands

*Tel/Fax:* 31-20-671.1818

*Email:* [greynet@inter.nl.net](mailto:greynet@inter.nl.net)

<http://www.konbib.nl/infolev/greynet>

# SIGLE

SYSTEM FOR INFORMATION ON GREY LITERATURE IN EUROPE

**The premier source for the latest advances  
in natural and social sciences, technology  
and medicine**

- ▶ **PROVIDES:** quick and easy access to European literature, not available through normal bookselling channels
- ▶ **ENSURES:** document/full text supply for each cited item
- ▶ **CONTAINS:** research and technical reports, preprints, dissertations, conference proceedings, etc.
- ▶ **ITS SIZE:** 565,000 records (Dec. 1998)
- ▶ **THE UPDATE FREQUENCY:** monthly

## **SIGLE is available:**

- ▶ **ONLINE:** **Blaise Information Services**  
<http://www.bl.uk>  
e-mail: [blaiseline-helpdesk@bl.uk](mailto:blaiseline-helpdesk@bl.uk)  
Tel.: +44 (0)1937 546430  
**STN International**  
<http://www.fiz-karlsruhe.de>  
e-mail: [hlpdeskk@fiz-karlsruhe.de](mailto:hlpdeskk@fiz-karlsruhe.de)  
Tel.: +49 (0)7247 / 808-555
- ▶ **INTERNET:** **STNEASY**  
<http://stneasy.fiz-karlsruhe.de>
- ▶ **CD-ROM:** **SilverPlatter Information Ltd**  
Tel.: +44 (0)181 995 8242  
Fax: +44 (0)181 995 5159

**SIGLE is a product of**



an association of leading information and document supply  
centres in 15 European countries.

For detailed information and document supply see

<http://www.konbib.nl/sigle>

## GREY LITERATURE: PLAUSIBLE APPLICATIONS FOR DISTANCE EDUCATION PRACTICES

Julia M. Gelfand  
University of California, Irvine  
[jgelfand@uci.edu](mailto:jgelfand@uci.edu)

Paper to be presented at the  
Fourth International Conference  
On Grey Literature  
Washington, DC  
October 4, 1999

**ABSTRACT:** With distance education among the most dynamic elements in post-secondary education today and with changes in emerging instruction technologies breaking frontiers, new interpretations of copyright and creative site licensing of information, grey literature has an equally new potential for use in these structures. The "Innovation Diffusion" common in distance education is a parallel structure to developments in electronic publishing and alternative ways to distribute and treat grey literature. Connectivity remains the major challenge as distance education courses are seeking materials and resources that they can ascertain will be available to their students via course supported software and a host of browsers with few restrictions. Grey literature, due to its wide scope and definition, and oftentimes unrestricted access, lends a significant use in this arena. There are mutual benefits to the author/editor/compiler, to the faculty, students, the host institution and that library, that may have some lasting impact in how distance education can transition from a dependent extension of traditional classroom based education environments to a widely independent distributor of information. This paper explores different methods of how grey literature can be more widely utilized in such applications and what challenges there are to promote grey literature in what appears to be both asynchronous as well as synchronous environments.

"The time is right for distance learning" begins the 1999 imprint, *McGraw-Hill Handbook of Distance Learning*.<sup>1</sup> What is implied here, is that as we near the millennium, it may be time to reduce the often harsh criticism that is levied against distance education programs and to see why alternative education is so attractive to a growing number of students and campus administrators and leaders. The information age has introduced new methods of delivering instructional materials and content to the student by creating nontraditional classroom experiences. These learning environments have changed rather dramatically as new and emerging technologies can now offer



instruction via a host of interactive channels. These changes create new opportunities and challenges for both student and institution of higher learning.

Distance education is a marketplace in which students are separated from their instructors by time, and/or space. The early history of the distance education movement was born with a dependence on correspondence education, in place since the late nineteenth century, where students received course materials via the mail, acquired materials, completed assignments and returned them to be graded again by postal service. Audiocassettes and broadcast mediums over television and satellite with add-ons such as video-cam recorded tapes, and finally, teleclasses involving the presentation of classes in specially designed classrooms, facilitated the next generation of technologies, followed by the introduction of computers. Classroom architecture was redesigned to accommodate and facilitate the new technology requirements.

When computers came along, in time they came with CD-ROMs, networks, potential for electronic mail, the Internet, mixed media, a vast, rapid emergence of online services using telephone lines and communication software. Thus, what has emerged is a vision of a networked learning environment that "seamlessly integrates voice, video, and data connections between and among learners, instructors, subject matter experts, virtual libraries, Internet resources and support service organizations. The distance learner is at the center of this network, connected with both real-time and non-real-time links to a wide variety of learning resources. The learning resources are themselves interconnected to form a mosaic of mutually enriching elements."<sup>2</sup>

As more of the general population becomes increasingly computer literate, the windows of opportunity open to invite new ways of doing business in education. E-



commerce suggests different connotations and meanings, but could easily define the new educational enterprises that have developed since barriers were reduced due to a networked environment. Distance education always served the adult learner, the student who was isolated from attending school in a traditional model and had to make alternative arrangements. It may have been because of geography and remoteness, economics, circumstance of needing to work and care for family that forced the issues of creating new models. Regardless, today those reasons are immaterial, because distance education is becoming a choice that extends beyond those considerations, and has proven itself as an alternative to traditional learning styles of being lectured to in increasingly overcrowded halls. Instead, it can offer a more flexible and individualized instruction package to students, with potential for more ways to interact with other students and the teaching faculty.

Computers are the most versatile components of distance education and provide the medium of communication and serve as the stepping stones of interactivity creating asynchronous environments. Software packages allow for institutions to customize online classrooms so students can:

- follow course schedule,
- obtain their lessons,
- ask questions of the teacher,
- share ideas with other students,
- participate in group and team assignments,
- request and transmit texts and graphics from content providers, and
- develop verbal and presentation skills by reading and writing, and engaging in creative measures.

Thus, a virtual classroom now generally offers:

- access to e-mail,
- threaded discussions,
- chat sessions,

- a whiteboard,
- shared applications,
- streaming video/audio and multimedia files,
- course and/or bibliographic management, links,
- interactive ROMs to interact with the course website.<sup>3</sup>

A recent study reported in the media and conducted at UCLA, says that the stress of staying up-to-date with technology affects more professors than traditional stresses such as teaching loads and publishing demands. The survey shows that 67% of professors are regularly stressed by keeping up with emerging technology, compared with 62% stressed by teaching loads, and 50% stressed by research or publishing pressures. Researchers say fear may be preventing professors from using new technology. Only 35% of professors use the Internet for research purposes, while 38% use technology to prepare presentations for classes. The survey results indicate that colleges should work to improve instructors' computer skills in order to meet the needs of students who have grown up using computers and are comfortable using new technologies.<sup>4</sup>

Many new concepts have been defined as well as the release of products associated with distance learning. They include "thoughtware" which refers to how "knowledge is so important, yet so underestimated and overlooked...that it is critical footing for the installation of new thoughtware in any organization. It's threatening, then disruptive, becomes provocative, then challenging and finally a true agent of change. Learning and development of new knowledge must become a part of every person's accountability."<sup>5</sup>

Some of the difficulties that have been well identified and articulated in scores of articles and books, are that institutions that played significant roles in teaching now must cope and adjust to new operational paradigms. These institutions include publishing and

libraries. Technology has also served these crucial pillars of education well, but challenged them in very complex ways. New relationships emerge with the concept of who is an author at the core of the exchange. Simply stated, the commercial product that has changed most significantly is the textbook. Textbook publishing has moreorless commanded the ultimatum, "publish online or perish."<sup>6</sup> In addition, servicing of the textbook has also dramatically changed, with electronic versions of library reserve services well in place with easy connections via Library online public catalogs, as well as the role of the campus bookstore, customized textbooks on demand, the course homepage, commercial lecture notes, and an array of other services and products.

The publishing and library communities have been very concerned about the impact of distance education. They have developed many different guidelines to direct the legal, social and ethical distribution of resources to students wherever they are, on campus, off-campus, in distance education or extended campus programs, or in the absence of a campus at all. Servicing course needs in distance education has become one of the fastest growing segments of academic libraries, with scores of newly defined positions for librarians where the emphasis is to provide content and services to the distance learner as seamlessly as to a local student. The Association of College and Research Libraries (ACRL), among the largest divisions within the American Library Association passed a set of revised guidelines for distance learning library services in July 1998.

This document carefully:

- defines all the players in the distance learning context;
- offers a set of philosophical assumptions about distance learning;

- lists how management of such organizations and institutions participating in distance education should consider administering library and information services;
- offers what functions library personnel associated with such a program should be engaged in;
- suggests the kind of facilities that lend themselves to the best blend of library resources and services;
- notes that thorough documentation should support all aspects of the efforts to provide adequate resources, access and information services to support all distance education measures.<sup>7</sup>

It is critical that such guidelines allow for flexibility and changes in the educational setting and promote that local applications should freely explore the best ways to accommodate faculty and student needs for efficient and effective information delivery.

So, where does grey literature enter into this discussion? If we think of grey literature as an adventure that can best be described as:

- 1) it uses knowledge and available data to confront tangible problems
- 2) it immerses the practitioner in real world situations
- 3) it often exists as an ill defined multidisciplinary field lacking clear boundaries,

then the future is clarified and one begins to know the concepts and methods of its construction. Grey literature has been known to enhance the service delivery of information by providing an alternative way of being distributed. It is a tool that builds human capital when used in locating other resources.

Distance education is also an adventure with sometimes a murky future; it has a risky undertaking, is a business venture with a remarkable and exciting experience to offer all who choose to participate. Can't the same be said for grey literature – the prospects are risky, and the journey is a remarkable and exciting treasure trove.



Distance education appears to flourish because students in a variety of settings can now have access to resources for which they have to go nowhere to get. The viability of the Internet as a place where resources are located makes the possibility of offering a wider variety of courses spanning the universe of information greater. With institutions entering into contracts that creatively negotiate the opportunities that libraries can promise for their affiliated students to gain access to materials they need to support their instructional needs, half the battle is won. The other half has to do with ascertaining that copyright compliance is practiced and this can be the more difficult part to achieve and oversee.

Today, grey literature is defined as information output that is not controlled by commercial publishing interests and where publishing is not the primary business activity of the authoring unit or organization. Thus, it is my contention, that grey literature has a natural home supporting distance learning initiatives. Since the span of distance education is now so encompassing and broad, there is even a greater sense of the kind of material that is appropriate to support the learning environment. Students will be conducting more independent research, spending time online and mixing media. As long as copyright becomes the focus and permissions are secured when needed and clearly stated, grey literature may well hold the tenets of intellectual property to serve the range of distance education. Higher education is expected to nourish the creativity, resourcefulness and independent spirit required to solve difficult social and scientific problems. It places an emphasis on independent learning, reasoning and growth.

The sources and examples of grey literature will never be sufficient to address all the demands of curriculum nor be discovered by those who need to use them, however

alternative electronic publishing readily and intensely fuel these needs. If grey literature continues to be well represented in the public domain, it increases the likelihood that it will be found, identified and used more seriously, thus will be cited and referenced in mainstream publications as well as in new forms of grey literature. This will include primary and secondary publishing output. The latter is likely to be more compatible with instructional materials. Examples of this may include interpretations of data sets and statistics, databases of images, concordances of notes and reference works, indexes to important but low-use or low impact works, and student produced work that has not experienced peer review. The latter is a bit scary, because it lacks authority control and contributes to the already disproportionate amount of material on the web with independent domain addresses.

Faculty seeking appropriate and relevant materials may steer students to such grey literature content if it does not require the potential expense, and planning ahead necessary to obtain the proper site license. Experience to date suggests that specialized course homepage has a strong future.

*The Report on Copyright and Digital Distance Education* issued by the U.S.

Copyright Office in May 1999, interprets how the Digital Millennium Copyright Act of 1998 (DMCA) is served by distance education. There are several parts to this report. For the sake of this paper, the most relevant parts are:

- Part I provides an overview and illustrates how current distance education practices operate.
- Part II is the introduction of current licensing practices in digital distance education and with a description of past and current problems and suggests trends.
- Part III describes the status of technologies relating to the delivery and protection of distance education materials.
- Part IV is an analysis of current copyright law and digital distance education.

- Part VI is the examination of whether the law should be revised and what the Copyright Office recommends at this juncture in time.<sup>8</sup>

The Report, hailed as welcoming by the higher education legal communities, responds very well to how the capabilities of digital technologies lend to expanding pedagogical possibilities within the classroom experience. One constant in teaching is to develop critical thinking skills. Encouraging students to seek and retrieve a variety of relevant materials that support their information query can do this. Usually consulting pathfinders, indexing tools or manipulating a variety of metasites and search engines does this. Students then must explore, understand, use, apply and evaluate these resources, and for some people, they may be logical and for others, irrelevant and ambiguous. Alistair Fraser writes, "In the past we relied on words, diagrams, equations and gesticulation to build those models (of how the world works) piece by piece in the minds of the students. We now have a new tool – not one that replaces the older ones, but one that greatly extends them: interactive computer visualization. Today, a teacher can build a pedagogical model, and both student and teacher can interact with it to explore the behavior of the system in a way inconceivable in earlier times."<sup>9</sup>

The increasingly stressful and problematic concerns in the distance education marketplace will not vanish; instead they may become a more complex set of issues, as choice grows and distance learning becomes even more widespread. With students taking degrees from institutions on foreign continents, the institutions of the Ivy League dipping their toes into this once forlorn land, mergers and affiliations with the British Open University, Virtual Universities springing up in every consortia and governed body, with the commercial and private sector well positioned to sell, serve and participate, the blurring of profit and non-profit entities, higher education establishing partnerships with



corporations and industry, it is not clear that we can even rightfully imagine what the full spectrum of distance education will be in three, five and ten years out.

Students will be granted degrees from institutions to which they have only visited online, and our netiquette should improve as we develop casual and long term relationships with those we only know as an e-mail handle. Grey literature will be there, coming up for air, taking a big breath and reinventing itself in a more civilized and communicative foray in as much of a technology neutral environment as possible. Grey literature will experience more company in its domain, with authors internal and external to the academic community wanting to contribute. The dilemmas of how to provide adequate and current enough bibliographic control will keep library staff actively concerned about how to compete with demands, and respond to legitimate and growing fears of unauthorized downstreaming of grey and not grey materials.

I can not help but be inspired by a statement in the conclusion of Albert Borgmann's book, *Holding on to Reality: The Nature of Information at the Turn of the Millennium*, when he writes, "Stories are the spaces wherein pictures and mementos come to life and coalesce into a coherent picture of the past and a hopeful vision of the future. Records in turn keep stories straight and lend them detail. Thus the culture of the word can card, spin, and knit the mass of technological information into a tapestry that is commensurate with reality."<sup>10</sup> My view suggests that is a good way as to how grey literature can be summarized. Efforts to create new learning communities, and to sustain and develop the better elements of contemporary distance education have a strong future as long as the technology can be supported for both asynchronous and synchronous learning environments.



<sup>1</sup> Alan G. Chute, Melody Thompson & Burton Hancock, *The McGraw-Hill Handbook of Distance Learning*. New York: McGraw-Hill, 1999, 1.

<sup>2</sup> *Ibid.*, 207.

<sup>3</sup> *Report on Copyright and Digital Distance Education*. Washington, DC: U.S. Copyright Office, May 1999. Available at <http://www.loc.gov/copyright/> 54-56.

<sup>4</sup> Angetta McQueen writing for the AP August 30, 1999 about a research study conducted by Linda Sax at the University of California, Los Angeles. No further citation available.

<sup>5</sup> J. Philip Kirby, *Thoughtware*. Portland, OR: Productivity Press, 1997, 130-131.

<sup>6</sup> Jeffrey R. Young, "A Pilot Project Adds Security to Electronic Textbook Publishing," *Chronicle of Higher Education*, August 6, 1999, A31.

<sup>7</sup> ACRL Guidelines for distance learning library services, final version, approved July 1998, available at <http://www.ala.org/acrl/guides/distlrn.htm>

<sup>8</sup> US Copyright Office, *Report on Copyright and Digital Distance Education*, Executive Summary, May, 1999, I-xxvi.

<sup>9</sup> Alistair B. Fraser, "Colleges Should Tap the Pedagogical Potential of the World-Wide Web," *Chronicle of Higher Education*, August 6, 1999, B8.

<sup>10</sup> Albert Borgmann, *Holding onto Reality: The Nature of Information at the Turn of the Millennium*. Chicago, IL: The University of Chicago Press, 1999, 231.

# SMAISMIRMILMEPOETALEUMIBUNENUGTTAUIRAS

## The internet & the socio-structural change of informal scientific communication

Helmut M. Artus

Social Science Information Centre (IZ)

### Introduction

To begin with: The *sub-title* of my paper is „*The internet & the socio-structural change of informal scientific communication*\*. Please do not urge me to read its *main title* which is not only even more complicated than the sub-title but is definitely unpronounceable and definitely not understandable. Nevertheless, there *is* sense and information in it, even information about an important empirical observation. I shall give the solution and explication in part 2.

My paper deals with the impact of the internet on scientific communication and publication and tries to find out which changes can be expected. At first sight, this seems rather clear and easy, since technical development seems to determine the future.

- So, *part 1* of my paper will deal with the *chances and possibilities opened up by technical development*.

On the other hand, we cannot expect that everything that is *possible* will also be put into *practice*. Not any *vision* is becoming *reality*, neither in politics nor in science. Between possibility and reality, there are things like *acceptance* or *refusal* which turn out to be really *social* processes, which means that they are *complicated*, *nonsensical* and *unpredictable* - at least as long as we did not understand the rules of the game.

- So, *part 2* of my paper has to do with *science as a social system, communication as a social process and scientists as social beings* - and as individual subjects persueing their own interests at least as much as the interests of science.

This part tries to learn from the history of science as well as from sociological theories explaining it. As we shall see, this sociological view can considerably change our expectations...

- ... as will prove *part 3* which tries to summarize the facts and arguments and make a prognosis of what will happen in the area of scientific communication. There, we shall have to discuss the *transition from virtual changes to real changes*.

### 1. Virtual changes

The internet has been changing our communicative customs in a rather dramatic way, and the end is not in sight by far, yet. Beyond, we can be sure that even the sector of publication - and above all that of grey literature - will undergo severe changes.

Just in order to illustrate this:

- My last five publications which came out this year, were presented in a printed and an internet version at the same time, three of them were additionally passed to about 2,000 addressees by an e-mailing list. The print version, by the way, is sent to 'only' 1,400 people.
- Since about two years, I have been receiving hundreds of e-mails per month coming in from individuals, but in most cases via mailing lists or discussion groups. Since that time, the number of mails I am writing my own has increased as well and is much above the number of letters I used to write.

Such developments have changed our behaviour in an interesting way. Since communication is much easier, we communicate *more, actively* (i. e. *writing*) as well as *passively* (i. e. *receiving* and - sometimes - even *reading*). Using e-mail or the internet, it has become possible to get information instantly or to receive an answer to your question within hours or even minutes. So why write a letter, look for an envelope, find out the zip code and postage for a remote country and get the stamps you need, and at last drop it into the mail box, if you can make things short and easy: just write a text and send it (usually without reading it again) to the other end of the world just by a mouseclick?!

So I think it is obvious that certain trends are inevitable:



- *Number of (electronic) communication:* There will be much more communication since e-mail and internet make things much easier, much faster and much cheaper.
- *'Density' and intensity of (electronic) communication:* The number of people participating in communicative or discursive connexions as well as the 'density' of communication will grow, i.e. the number of contributors as well as the number of contributions. The net of questions and answers and statements will get closer and more intense. The reasons - besides the ones already mentioned - are that the anonymity of such connexions diminishes inhibitions and encourages spontaneity, so that a lot of people do participate at all or more often.
- *Acceleration of (electronic) communication:* Communication will become much faster which does not only mean that it is technically transmitted in a much shorter time, but as a result, discourses will be transacted in a shorter time. Either you participate in a topical discussion within a few days - or you are late and the topics discussed have completely changed meanwhile. Ironically just the immense saving of time causes an enormous time pressure. So it could be asked whether - as far as communication is concerned - the internet really is a time saver or maybe rather a time consumer! In fact, it saves time - but we are not free to make an optional usage of it because by that same time saving process it imposes time pressure. I think, most people talking about time saving are not aware of this dark side of the net...

Similar trends will occur in publishing:

- *Increase of (electronic) publications:* We can be sure that the number of (electronic) publications of any kind will dramatically increase. At least in the industrialized countries everyone has his own computer, mostly with an internet access. So anybody can create his own homepage and bring his non-commercial one man publishing house into the world wide web so that the whole world has a chance to download and read his papers.<sup>1</sup>
- *Acceleration of electronic publication:* We can count on an acceleration of (a part of) publications as well, since the internet allows the avoidance of the time consuming (and - as seen from the author's point of view - risky) referee system deciding about the acceptance or refusal of his paper. And at the same time, the gap between the technical processes of producing and publishing (printing) is removed since there is no considerable expense to prepare a text file for 'publication' in the net.

<sup>1</sup> A few days ago I read about the representative of a famous German publishing house, saying that the 'end of the Gutenberg galaxy is not yet in sight': Publishers still earn their money with printed

- *Extension of our understanding of literature:* Short papers for which there was no chance of publication in the past could get a chance now to be published in the internet. Thus, a new genre could be created, short and very grey treatises which used to be kept under lock and key in former days or were only communicated in small circles with restricted access. So the size of a paper ceases to be a criterion for publication. It is no longer the number of pages that decides whether a paper is (or can be) 'literature' or 'nothing'.
- At the same time (and for the same reasons as mentioned with communication), the inhibition of publishing papers will be reduced. This will lead to an increasing number of people publishing at all as well as of the average number of their publications.

If we add up one and one, we must draw the conclusion that boundaries are shifting, be it those between communication and publication which used to be clearly separated facts, be it between grey and white literature, be it between grey literature and what I would like to call *dirty-grey literature* (dirty grey, not dirty literature!), by which I mean: communication which is written and not private, but is still beneath the level of grey literature, e.g. has no institutional background, no formal 'bibliographic' (or *internet-ic*?) properties etc. So the whole sector of scientific communication - be it private, informal or formal - is in motion, a proper revolution seems to be in full swing. It seems, that there will be a competition between (printed) *grey literature* (GL) and *electronic literature* (EL). If we only gaze at the technical facilities and possibilities, the fight is already decided: *EL beats GL by technical knockout!* But don't abandon hope! The next chapter deals with social facts, and that can change a lot! So be of good cheer!

As a consequence we could expect that the greatest loser of the *internet revolution* is exactly the kind of literature we love most, i.e. grey literature. So perhaps we better should write an epitaph on a species of literature becoming extinct.

Functions that used to be fulfilled by grey literature - especially the quick and uncontrolled, informal dissemination of facts and figures, of unfinished ideas and unaccomplished interpretations - can be fulfilled by the internet as well and in many cases even better, much better.

Nevertheless, I do not believe in the sudden death of grey literature but quite in contrary count on a glorious resurrection of that kind of literature, although in electronic shape. I believe that this *come back* will be really enforced by the processes happening in the internet. We shall see - in parts 2 and 3 of this paper.



## 2. A sociological view on science & scientific communication

Let me now come to the solution of the riddle given in the title of my paper.

### a) A historical episode

As I admitted before, the title is definitely *unpronounceable* and not understandable. Nevertheless, there *is* sense and information in it, even information about an important empirical observation.

The tape-worm-like word is a so-called *anagram*, which means that the author - a certain *Galileo Galilei* - has taken a whole sentence indicating a brandnew observation of his and arranged the letters in a completely different order thus definitely excluding any understandability.

The original sentence - in Latin, of course - reads

*„Altissimum planetam tergeminum observavi“*

which means:

*“I observed the highest planet (=Saturn) in threefold form.”*

I think, even this text needs another translation to be understandable. *Galileo* wanted to express that he had discovered *two moons* of Saturn. Actually, he had observed the *ring system* surrounding the planet. *Kepler*, the contemporary genius and (at least in *Galileo's* eyes) rival tried weeks and weeks to find out what *Galileo* could have expressed by his anagram, but failed.<sup>2</sup>

What happened here?

To understand this historical episode, which took place in the year 1610, we have to keep in mind that the reward system of science is somewhat different from that in business or in other societal sub-systems. There is neither *money* as a reward nor something like *intellectual property*, not even for the one who made a discovery. There is only *prestige and esteem* - but only for the one who is first! Only priority counts. There are no silver medals in the race of scientific priority. Either you win gold or you lose.

In such a system the indispensable interest of a scientist must be to be the first - and to keep a safe lead over all competitors. Actually, there is a remarkable contradiction (or at least something very close to a contradiction) concluding from the scientific norms and rules: If a scientist makes a discovery, he is first, but he can only earn the grapes of his

<sup>2</sup> Actually, he found a solution which he himself scornfully called a 'semi-barbaric Latin verse' and which, besides, was wrong: *„Salve umbistineum Martia pros.“* - „Salute, glowing twin, sprout of Mars.“ He thought *Galileo* had discovered a moon of Mars.



performance if he publishes his findings. Doing so, however, he gives his rivals the chance to surpass him. So by publishing his discovery he imposes a new stress on himself and can no longer be sure to keep the lead.

This was *Galileo's* situation, but he found a very interesting way of securing his priority without, however, allowing his rivals - above all *John Kepler* - to derive any profit from it. His idea was to 'publish' the completely un-understandable anagram.<sup>3</sup> By decoding his message, he could at any time prove his priority, and by not explaining its meaning he could be safe from *Kepler* and others.

### **b) The "storybook version" of science**

This is much more than just a historical episode but leads us straight into the social life of science, into the very human behaviour of scientists as human beings, and into the problematic and sometimes contradictory connexion between science as a *cognitive* and science as a *social* system. As *Galileo's* behaviour proves, there can be reasons for retaining information and for not or not really publishing them. The famous episode is a good example for certain contradictions between personal (subjective) interests of scientists as individuals and social norms of science as an intellectual connexion if not even for contradictions within the system of scientific norms itself.

I shall try to explain this in a few words.

In general, science is depicted as a *cognitive connexion*, i.e. an arrangement aiming at knowledge only. Each and everything happening there has a scientific, i.e. a cognitive, intellectual or theoretical meaning, individuals or groups acting within this system are subject to the idea of creating, testing or communicating scientific knowledge, of producing progress and so on. Although human beings, scientists do nothing than execute the rules of appropriate and methodologically correct research. As long as they act within the social system of science they guide their whole action along scientific knowledge. Actually, there are deviances from the rules of appropriate professional conduct, but these are just individual deviations which are sanctioned by the scientific community or even by juridical courts.

<sup>3</sup> Of course, *Galileo* did not formally *publish* the anagram expressing his discovery, but gave it to the *Toscanian* ambassador at *Prague* (*Kepler* was astronomer at the Emperor's court at *Prague* at that time). Three long month later *Galileo* explained the anagram's meaning - but not to *Kepler* or any other scientist, but to Emperor *Rudolph*; p. 382 in: *Arthur Koestler: Die Schlafwandler. Die Entstehungsgeschichte unserer Welterkenntnis* (first published in English: *The Sleepwalkers*, 1959), Suhrkamp: Frankfurt/M. 1980

What I have described here is what *Blaise Cronin* has called the '*storybook version of science*'. Scientific fairy tales. But nevertheless, this is the image and self-image of science, and even the sociology of science who should know better (and look at its object much more accurately, with skepticism and sense for realism), - even this sociology of science often fails to look at science as *it is* and instead describes science as *it should be*.

### c) A realist version of science

From a sociological (or simply *realist*) point of view this is sheer nonsense. From its very beginning, science has been a *social* connexion realised by human individuals, acting together (or against each other) in social groups, under social aims, values and norms. If we look at scientists, we see real social subjects, i.e.: people who earn their living by doing their jobs like bakers or engineers or shoemakers. So job and money are inseparable parts of science as a (really existing) social system and of a scientist's life. And scientists - as really existing people, not as bloodless abstractions as presented in the *storybook version* - , scientists are real human beings with individual interests and preferences and convictions and goals etc. of any kind - scientific as well as political, religious, philosophical, sexual, social and so on. As real human beings, they also have their characters and feelings or, in other words, a *psyche*. They are as noble, decent and generous or mean, evil and envious as any other people, they are jealous or cordial, they are intelligent or - sometimes - stupid, they are healthy or mentally ill as any other people. And last but not least they are egoistic, selfish, and vain. They do not only want money and awards, prestige and esteem, but also admiration, not to talk about sex. (But this is another story, that you cannot find in written descriptions but only in the *oral history* of science .)

If we accept this - and I think there is not one reason for denying it -, if we accept this we cannot exclude really human features from science but have to admit that scientists - as any other human beings - act in a way that is a mixture between professional rules and norms on one side and their individual characters and interests and psychological predispositions on the other. If you do not accept this, please remember the *Galileo* episode giving a very good example for this mixture. It does not only prove that a scientific genius as *Galileo* can have a rather bad character (which, to my knowledge, he had), but that individual features and interests can interfere with the scientific process so that science cannot be conceptualised without regard to the scientist as a real human.



#### d) Some remarks about social inequality in science

One should assume that scientists would appreciate any innovation promising to improve scientific communication and the dissemination of information and knowledge, such as the internet. But as we have seen, there are reasons for not communicating, for not revealing information, for placing personal interests above the interests of science and knowledge. These reasons can be cognitive or social, they can be compatible with appropriate scientific conduct and they can be very personal, selfish and contradict any scientific norm and value.<sup>4</sup>

The role of the scientist as an individual and as a really social being is one side of the problem. The other and much more important side is that modern science in the course of several centuries has been developing a comprehensive set of goals, values, norms and sanctions as well as social structures and of functions that have to be fulfilled, if science should work appropriately. Thus, science as a *cognitive connexion* is based on science as a *social connexion*. The social system deserves its goals and values to the *cognitive* definition of science, but on the other side, the cognitive connexion can be realised if and only if the social system works - and only in as far as this social basis works.

At first sight, it might be self-evident, that science is a deeply democratic domain, where equals interact in an ideal way. Actually, however, *the idea of science requires differences*: There must be something like social stratification, of inclusion & exclusion, of *social inequality*. If *prestige* and *acknowledgement* are the valuta of science and its reward system, as I have explained above, then all factors providing such necessary social differences are necessary for the functioning of science:

- There are, for instance, *competition and rivalry* for priority and acknowledgement; these are social processes, but their indispensable function is a cognitive one.
- Not publishing itself grants prestige, but publishing in a journal or a publishing house with high reputation. And the most reputed journal or publishing house is the one with the highest rate of refusals or with the highest barrier to access. The greater the difficulties, the higher the prestige for the happy few who succeed and whose articles are accepted.

<sup>4</sup> During the last decades, commercial reasons have increased. *Science based industry*, I presume, cannot follow exactly the same rules as basic research, and the same holds for research obliged to the sector of defense. As we know, considerable parts of e.g. informatics have not been publishing in the 'official' (open) publication system of science for years and years, but in a parallel one hermetically closed against the rest of the world. Here we find another version of *Galileo's public-secret-dilemma*: Scientists doing top secret research work can *publish* (!) their results - but only within a publication system which as a whole is a secret itself. So *keeping a secret and publishing it at the same time is possible*. Of course, this is a perversion of science.



This is not because scientists are secret masochists but because these are fundamental differences that make science work.

It is self-evident that those who have undergone all the problems, troubles and toils of the long and intricate way to the top, do not have the slightest interest in changing the system to whom they owe their prestige. This may be considered to be a more *psychological* and subjective attitude, but it isn't. The mechanisms providing the social inequality of science are basic. Without them, science would give up the idea of elites (and the difference between the elite and the scientific crowd, between the genius and the simple and unimportant researcher etc.). There is nothing else that could provide a substitute. Then, science would turn out to be a social system like any other without anything special and, above all, without any chance to claim a superiority of its special kind of knowledge as compared to any other.

So we can start from the assumption that these functions which are of fundamental importance are subject to a *meta-stabilization* warranting that their shape may change, that their outer appearance may differ from time to time, but that *if* they are replaced, they are replaced by other 'phenomena' fulfilling exactly the same functions. Anything else would mean giving up fundamental principles of science.

So despite any possibilities offered by the internet, we can expect that this "functional nucleus" of science, warranting social inequality of a certain kind, will not be touched, diminished or even destroyed or replaced by something else. So we can expect that the *appearance* of scientific communication and publication in the age of the internet could change, but at the same time we can expect that on the *functional* level (where all important things happen) *only little will change - if anything at all*.

### 3. Real changes: What do we have to expect?

Now that the sociological questions are somewhat clear, we can try to draw the conclusions. So, let me now come to the third and last part of my paper and to the question: *What do we have to expect?*

Let us at first ask what is going to stay although it may change its outer appearance:

There is no doubt that a lot of communication, grey and even white literature will find its way into the internet in the near future. I would not argue about that. But my *thesis* is that despite any changes caused by this new technical medium, there will be only little really *substantial* change. Scientists will *use* the internet but will not *subjugate* to it. They will by no means change their customs, their patterns of behaviour or even the social structures of

science or scientific communication. So the changes to be expected, I think, will be much less dramatic than they seem (or threaten) to be.

After all I have said, I think the following points are clear:

- The fundamental *difference between uncontrolled (public) and controlled communication* which is restricted to small groups and personally known participants will remain. It is an illusion to expect that the leading scientists and experts would open their communication channels to anybody.
- The same holds for the *difference between controlled and uncontrolled access to publication media*. Of course, anybody can create his own homepage and offer his private publishing house; in so far the internet will change a lot and provide more democracy. But not only *prestige* but also other scientists' *interest* (!) will now as ever be related to media with strictly controlled access. So the mere mass of additional electronic publications will neither question the boundary between restricted and unrestricted publication, nor will it change the course of the publication process. The transition from print to electronic publishing will not change anything of such a fundamental importance as the *social stratification of science*, i.e. the difference between top scientists and unimportant project researchers (and all qualitative levels in between).
- Also, the difference between *preliminary and final publications* will not be questioned, although the *channels* or media of information could change. As long as prestige has to do with priority, it is necessary that scientists can publish their findings before they are able to give their publication the final shape and present it in perfection. I am sure that they will find new and surprising ways of using the internet for this very purpose. But I am sure as well that the basic difference between such pre-information just in order to save a priority and the formal publication will remain the same, even if - at the surface - things look different. What was important for Galileo will be important for his colleagues in the internet age as well!

At first sight, it may look seductive to publish without any restriction, without a peer review system, without the risk of being refused by the editor. But this argument would hold for grey literature as well. Why did grey literature not replace formal publication? For the same reasons, we could ask now: Why will the unrestricted internet site not replace other publication forms which may be electronic but will be strictly controlled in any case? The reasons are the same in both cases: Neither grey literature nor the free internet site can create social inequality, which - as we have seen - is an indispensable function within the social system of science.



So let us ask the second question: *What will change?*

- Without any doubt, a lot of communication and publication is going to be electronic. The advantage of the internet is too obvious to waive its possibilities. Some of them, I have already described above: Communication will grow in number and become closer and more intense.
- A certain type of information, communication and publication will go over from print to the internet, notably *information with trivial contents*. There is a lot of information which is necessary for scientists although it is not scientific at all. Take for instance information about institutes or research projects, about the controlling gremia of research institutes, listings of approaching international conferences, brochures with new publications and so on. In the past, most of such information was published in form of prospects, in the future, they will be disseminated via internet pages. Trivial communication will use e-mail (as it already does).
- One of the most interesting and important things is not to stay but has still to develop. The internet is not only a medium for arbitrary usage, but a kind of *complex service center*. In a fascinating way it could link the different parts of the information sector which belong together but for historical reasons have undergone separate developments. These are commercial publishers, research institutes producing grey literature, libraries keeping literature in stock and lending it, book sellers and - last but not least - database producers and information brokers. Using the internet it could be (or better: *become*) possible to make an online database search, then - within the same online session - order a copy of an article from a library and download (for money) another one from the commercial publisher's server, order a book from a book seller and send an e-mail to the research team that has published some grey material of interest. This, I think, is what many scientists have been dreaming of for generations.

But what about *grey literature*?

- A lot of grey literature as well as of formal publications will be published electronically. But despite the transition to the internet, the *difference between grey and white literature will be maintained*. The financial interests of commercial publishers are as important as the scientists' interest in the *difference between refereed and non-refereed publication*. Of course, many new publications of any kind and quality will be accessible via internet - but be sure that a cash-box will be installed if the information is really worth reading or downloading. Trivial, unimportant, uninteresting and unrefereed literature will perhaps be free, but *anything of quality and interest will have its price*. The mode of account will of



course be electronic. So the invoice may be only *virtual*, but the money you have to pay is damned *real*!

- As pointed out above, the boundaries between communication and publication will be contested. Notably at the point where (at the time being) grey literature has a more *communicative* intention and aims at *feedback* more than at informing colleagues. - notably at this point, the demarcation between communication and publication will become blurred, texts will be brought into a *discursive* shape rather than being given formal preliminaries etc., so that they can be correctly cited. Such a development, by the way, should be welcomed since communication is one of grey literature's major functions. And if we keep in mind, that science as a whole is a discursive connection, we should be glad that the (to a certain degree) artificial boundaries between communication and publication are pulled down at least at one point. (Nevertheless, even the contributors of such communicative non-papers will take care that their personal performance can be identified and honoured. I think it was the famous sociologist of science *Robert K. Merton* who said that mere *participation* in the game of science was enough appreciation, so the authors should remain *anonymous*. But of course he knew very well that this was just a fiction since scientists are real human individuals with real human interests and vanity.)
- As mentioned before, grey literature will not be replaced by anything else. But on the other side, the producers of grey literature have to expect the most severe problems. More than any other *genre*, grey literature is threatened by an immense number of completely unrefereed, uncontrolled and undiscussed papers from '*one-man-institutes*' and '*one-man-publishing-houses*' - consisting of 1 person, 1 computer, 1 internet interface and at least 1 paper to be published. So if grey literature does not want to be confused with an uncountable number of (more or less) worthless or at least dubious papers, thus losing its reputation as a serious scientific information medium, then it has to draw consequences.

Ironically, the future of grey literature is not more freedom and less control, but quite in contrary a sort of *new institutionalization*. Grey literature, notably *electronic* grey literature - which, of course, is disseminated via internet -, *must* care for a clear and unequivocal demarcation against a cosmos of anything that is beneath its level. This, I think, is the *price for survival*. Grey literature (in our sense) must not be refereed (thus losing its identity) but has to get something like a *brand* or a *trademark* proving that its origin is a respectable institution or project and making clear that we do not have to do with "*dirty grey literature*" or "*private homepage literature*" with free access to each & every one.

**Conclusion: G. L. is here to stay**

The internet has been changing all our communication and publication habits, that is true. But it is true, either, that also in the internet *man* is measure of everything.

Lots of literature will be distributed via internet. We can expect some substantial changes, but only in a few cases. In most cases, however, the internet will only be used as a new technical medium for selling old wine in new bottles, without changing the substance of its contents, the functions or the publication and (partly) communication system of science.

And so, after all, I think that there is no reason to worry about grey literature's future.

## GREY LITERATURE AND LIBRARY AND INFORMATION STUDIES (LIS): A GLOBAL PERSPECTIVE

*Prof. L. O. Aina*

*Department of Library and Information Studies*

*University of Botswana*

*P/Bag 0022*

*Gaborone, Botswana*

*Tel: 367-359490*

*Fax: 267-585098*

*Email: ainalo@noka.ub.bw*

### **Abstract**

*A lot of studies on grey literature abound in the literature. Most of the studies, however, relate to disciplines other than LIS. The main objectives of this study are to determine the focus and trends of grey literature publications. Other objectives include determining the proportion of grey literature published that focuses on national, regional and global issues, identifying the various formats of grey literature prevalent in LIS, and the major sources used by information professionals in publishing their grey literature research. In order to carry out the study, an analysis of grey literature publications as recorded in Library and Information Science Abstracts (LISA) from 1993 to 1997 was carried out. During the period considered, using the free text option, 179 publications had anything to do with grey literature but only 101 or 56.4% were focussed on grey literature, thus only 101 abstracts were used as data sets for this study. The findings revealed that the major areas of concentration during the period considered were collection (46.4 %), information technology (18.2%), bibliographic control (17.3%) and characteristics of grey literature (13.9%). A large proportion of the researchers focussed on issues affecting Europe (39.6%). This is followed by global issues which constituted (33.7%). The majority of the research findings were published as conference proceeding (63.1%), Journal, as a medium of publication constituted (30.1%). Other media used were reports and books. International Conference on Grey Literature Proceedings constituted the major source for reporting grey literature as it accounted for (57.3%). The study recommends that the International Journal of Grey Literature recently launched should be given maximum support so as to serve a major medium for publishing research findings in grey literature. Other recommendations include the need for more research in grey literature in the other continents especially Asia where less than 5% of the grey literature published emanated from Asia, the need to research into other aspects of grey literature besides collection, characteristics and bibliographic control, as these three areas constituted more than 80% of the papers published.*

### **Introduction**

At the third International Conference on Grey literature held in Luxembourg, in 1997, a consensus definition of grey literature ( GL ) was adopted. This comprehensive definition states that "Grey literature is that which is produced at all levels by government, academia, business and industries, both in print and electronic



formats but which is not controlled by commercial publishing interests and where publishing is not the primary activity of the organisation" (Farace, 1997). In spite of the fact that it is not commercially available, it is generally recognised that it is the first port of call, when reporting important findings, be it research, corporate or even individual activities. It is therefore not surprising that GL comes in the form of policy documents, corporate reports, conference proceedings, theses and dissertations, feasibility studies, working papers etc. There is no doubt that grey literature is valuable in research activity, policy and decision making. In some disciplines, it constitutes a sizeable proportion of the literature. According to Chilag (1982) much of knowledge and information will never appear in any other form except grey literature. A study in grey literature use on agriculture revealed that as much as 14% of grey literature was used for research in agriculture in Nigeria (Aina, 1987). The past three international conferences on grey literature have brought out clearly the need to manage grey literature just as the white literature. Given the fact that the paradigm shift is from ownership to access, many research workers have emphasised the need to market, advertise and disseminate grey literature so as to provide greater access, because of its importance in the generation of new knowledge.

Library and information workers constitute one of the important stakeholders in making grey literature accessible to all who those who need it. Many studies have been carried out to show the use and importance of grey literature in many disciplines other than library and information studies (LIS). This present paper aims to find out how much grey literature is used in reporting research and corporate activities in library and information studies. The main objectives of this study are to determine the focus and trends of grey literature publications in library and information studies. Other objectives include determining the relative proportion of grey literature published, that focuses on national, regional and global issues; identifying the various formats of grey literature prevalent in LIS, and the major sources used by information professionals in publishing their grey literature research.

There are a variety of ways for studying the use patterns of grey literature by professionals in any discipline. One way of carrying out the study is to survey researchers, using questionnaire and interview as survey instruments for collecting the data. This will involve seeking their opinions on the frequency of use of grey literature as an important medium for reporting research findings.

Another alternative is to compile a comprehensive bibliography of the writings of professionals in any discipline and then analyse it bibliometrically to find out the proportion of what constitutes grey literature vis-a-vis white literature. A more direct way, however, is to use an existing data base on library and information studies

research, which can then be analysed. This is the method adopted for this study.

In carrying out this study Library and Information Science Abstract (LISA) database was used as the source data. This is based on the assumption that no abstracting agency will abstract a worthless document given the time and efforts needed in writing an abstract for a document. Thus, any important document published on grey literature will be covered by LISA.

LISA is the principal abstracting agency for abstracting documents in the information profession throughout the world, and it was the natural choice for this study. The assumption of this study is that LISA accurately represents the research strength of every country of the world in the information profession and whatever is covered during the period considered accurately reflects the output of each nation. The electronic version (LISA Plus ) 1998 was used in identifying all the literature pertinent to all the various countries and regions of the world. Using the search facilities of LISA Plus, the free text and country of publication fields were searched for all countries in the world. The countries were then grouped into regions. For each paper covered by LISA, the subject of research, country/region to which the research was addressed, the format of publication, and the name of the journal, if the format of the paper was a journal, were recorded on a cataloguing card. In order to carry out the study, a survey of grey literature publications as recorded in Library and Information Science Abstracts (LISA) from 1993 to 1997 was used.

## Results and Discussion

An analysis of LISA during the period covered, revealed that there were 179 publications that mentioned grey literature, although only 101 publications were concentrated on grey literature. Thus, 101 publications were used as the database for this study.

Table 1 provides the geographical coverage of grey literature for the different regions of the world. The various countries were grouped into four regions - namely Africa, Asia, Europe and America were identified. Findings reported under global reflect the documents that address issues that go beyond national and regional boundaries. From the table, it is obvious that Europe (39.6%) is the single largest region that has made more contribution on GL than any other region. This is not surprising given the fact that research on grey literature use has been extensive in this region. European Association for Grey Literature Exploitation (EAGLE) has been the principal agency for promoting the use of grey literature in the region. A sizeable proportion of the papers focussed on global issues (33.7%). It is also significant that more papers were recorded in 1994 and 1996 which tally with years



the GL proceedings were published.

Table 1 : Geographical Coverage of Grey Literature Publications (1993-1997)

Distribution	1993	1994	1995	1996	1997	TOTAL	%
Europe 39.6	5	16	7	10	2	40	
Global 33.7	1	12	3	8	10	34	
Africa 13.9	1	6	1	6	-	14	
America 7.9	-	3	1	4	-	8	
Asia 4.9	-	3	-	2	-	5	
<b>TOTAL</b> <b>100.0</b>	<b>7</b>	<b>40</b>	<b>12</b>	<b>30</b>	<b>12</b>	<b>101</b>	

Table 2 Formats of Grey literature Publications (N=103)

Format	Number	%
Proceedings	65	63.1
Journal	31	30.1
Report	4	3.9
Book	3	2.9
<b>TOTAL</b>	<b>103</b>	<b>100.0</b>

Table 2 reveals that there were some publications that appeared in more than one format. Some papers which have earlier appeared as proceedings and which were



later published as journal articles, hence the number of publications is two more than 101. However, from the table it is obvious that most of the GL publications in library and information studies appeared as proceedings (63.1%). A further analysis of the specific sources used for reporting GL as revealed in table 3 shows that 75.7% of research findings in GL were reported only in seven specific sources and that the *Proceedings of the International Conference on Grey Literature* published in 1994 and 1996 by the Grey Literature Service Network, the Netherlands, which is the major sponsor of the biennial conference on grey literature was the major medium for reporting grey literature. Majority of the publications appeared in the the International Conference on Grey Literature Proceedings ( 57.3%). Journals (30.1%) as format for reporting grey literature constitute a significant proportion, although the publications seem to appear in a variety of journals. Twenty-one journals were responsible for publishing 31 articles. With the exception of *Publishing Research Quarterly* (5), *Bolletin AIB* (4) *Serial Review* (2), *Documentation et Bibliothèques* (2), *Interlending and Document Supply* (2), only one article each appeared in the remaining 16 journals.

Table 3: Major Specific Sources for grey Literature (N= 103)

Sources	Number	%
Proceedings of the International Conference on Grey Literature	59	57.3
Publishing Research Quarterly	5	4.9
Bolletin AIB	4	3.9
IATUL proceedings	4	3.9
Interlending and Document Supply	2	1.9
Documentation et. Bibliothèques	2	1.9
Serial Review	2	1.9
<b>Total</b>	<b>78</b>	<b>75.7</b>

Table 4 shows the distribution of the subject content of GL publications during the period considered. The total number of publications is 110 as opposed to 101. This is because some publications cover more than one subject area. The table shows that researchers in LIS focussed more on collection of GL more than any other area. This is not surprising given the fact that a major function of the information professionals is to collect GL. Other significant subject areas focussed on include information technology which is expected to be the main plank for making GL accessible. Of course bibliographic control continues to be a major area of interest to information professionals since there is a need to move away from ownership to the provision of access. The emphasis on characteristics of GL is not misplaced because it is perhaps important to identify its features so as to be able to define GL appropriately, hence the consensus definition arrived at the 1997 International Conference on Grey Literature.

Table 4: Distribution of Subject content of Grey Literature (N= 110)

Subject	Number	%
Collection	51	46.4
IT	20	18.2
Bibliographic Control	19	17.3
Characteristics	14	12.7
Co-operation	2	1.8
Communication	2	1.8
Marketing	1	0.9
Copyright	1	0.9
<b>TOTAL</b>	<b>110</b>	<b>100.0</b>

### Conclusion and Recommendations

From the study it is apparent that there are few outlets for reporting research

findings in LIS. Unlike in most other fields where journal format constitutes the principal medium for reporting research findings, the opposite is the case with GL where proceedings of conferences provide the medium for reporting GL. This further makes access to GL even within LIS more difficult. The study also confirms the leading role of GreyNet (Grey Literature Network Service), Netherlands for being the principal agency for promoting research in GL. The study also reflects the need for the consensus definition of GL as adopted at the last GL conference. This is because most of the areas concentrated on, during the period considered, were collection and characteristics of GL. However, with the adopted definition, it is expected that the focus on these areas will shift to access. Europe is the most active region for the promotion of GL as already shown in table 1. Surprisingly Asia contributed less than 5% to GL research during the period considered.

The recently launched *International Journal of Grey Literature* has really come at an auspicious time, it is hoped that the journal will be supported by researchers in the field. Given the fact that access rather than ownership, is the norm, now it is expected a lot of research will now focus on how information technology can be utilised to ensure that GL becomes more accessible. It is hoped that other regions will focus more on GL just as Europe.

#### References.

- Aina, L.O. (1987) Grey literature and Agricultural Research in Nigeria. *IAALD Quarterly Bulletin*, 32 (1)47-50.
- Chilag, J. (1982) Non-Conventional Literature in agriculture-An Overview. *IAALD Quarterly Bulletin*, 27 (1) 2- 7.
- Farace, Dominic (1997) Third International Conference on Grey literature held in Luxembourg, 13-14 November 1997.



## Perspectives on Electronic Theses and Dissertations

Gail McMillan

University Libraries, Virginia Polytechnic Institute and State University

### INTRODUCTION

This paper provides three perspectives on electronic theses and dissertations (ETDs): the student's, the faculty member's, and the librarian's. These might also be described as immediate, near term, and long term perspectives. While this paper describes largely recent events, it also draws on the ten year history leading to the Networked Digital Library of Theses and Dissertations (NDLTD), especially the effects of the ETD requirement at Virginia Tech since January 1, 1997. A surprise to some, a revelation to others, but it is a fact of life now for the students, faculty, and library at Virginia Polytechnic Institute and State University (Virginia Tech). As a result, we have gathered a body of information, factual and anecdotal, that describes the impact of ETDs on many cultures both within and outside the academy. This paper presents the ETD initiative from these perspectives.

### STUDENTS' PERSPECTIVES

When graduate students at Virginia Tech submit their ETDs, they are asked to respond to a nine-question online survey. The following summarizes the surveys received from 455 graduate student authors submitting their ETDs between February and mid-August, 1999.

When they needed help with their questions about ETDs, 60% of our students sought personal assistance, turning to their friends, committee members, or staff at the library's New Media Center (NMC). The most frequently consulted source of personal assistance to authors' questions about ETDs was their friends (42.6%). Graduate students' consulted their committees nearly 8% of the time, and the NMC staff worked with nearly 10% of our ETD authors. It is the only computer lab that has staff trained and assigned to help its users. Other open computer labs on campus have computers and software, but staff monitor the facility and are not assigned to assist users.

When they wanted answers to their questions about ETDs, three-fourths of our students also turned to the Web. 74.9% said they consulted the VT Web and a few (1.1%) went elsewhere on the Internet. We asked authors who used the VT Web (<http://etd.vt.edu>) how helpful they found it, and 72.2% found it helpful or extremely helpful. Only 1.6% said it was extremely unhelpful and 6% said "unhelpful." 13.9%

found the Web neither helpful nor unhelpful. Staff working with the Digital Library and Archives are reorganizing and updating the VT Web so it will continue to be useful.

Another source of helpful information was attendance at the ETD workshops organized by the Graduate School. Presented twice a semester and once during each summer session, the workshops are held at a variety of locations and at times meant to appeal to students. These workshops have two goals: to introduce students (and faculty who may attend) to the ETD process, and to provide help and advice for preparing ETDs. The workshops are presented in two segments: the first portion is one for the "novice" and the second is for those with more advanced skill levels. The former emphasizes some typical word processing steps, such as using page breaks and how to automatically insert page numbers, as well as preparing tables of contents and integrating portrait and landscape page layouts in the same document. The instructor also covers the steps from the word-processed document to PDF (portable document format) and submitting an ETD through the Web to the Graduate School.

The ETD workshops, however, were not as successful as the Web site in providing graduate students with assistance, because only 21% said the workshop was a source of answers to their questions about ETDs. Of those who said they attended the workshops, over half, 56.9%, found them helpful or extremely helpful. A large percentage, 34.1%, was neutral on this issue, while less than ten percent did not find the workshops helpful.

Whether they got help through personal contact, the workshops, the Web, or elsewhere, over half (256) of our graduate student survey respondents were pleasantly surprised that turning their theses and dissertations into portable document format (PDF) files and submitting them electronically was easier than they expected. Over one fourth, 28.7% reported that creating PDFs were somewhat less difficult, and 27.6% found PDF conversion much less difficult than they had expected. Unfortunately, we still had nearly one-fourth of our students, 106 out of 453, who found PDF more difficult than they had anticipated. "Somewhat more difficult" was the response from 15.7%, and 7.7% said it was "much more difficult." Based on discussions with New Media Center staff, the ETD workshop instructor, and the Dean of the Graduate School, it appears that older graduate students have less experience with computer technology and, therefore, encounter more problems converting their



theses and dissertations to ETDs. Others that have difficulty are, not surprisingly, those who wait until the last minute and find they lack the time to learn to use the technology. One-fifth were not surprised one way or another by the conversion to PDF process.

The results from our graduate student authors were similar when surveyed about the ETD submission process. Over half, 56.5%, found it less difficult than they expected, and nearly one-fourth found it more difficult. As with converting files to PDF, one-fifth found it neither more nor less difficult to submit their ETDs than they had expected.

Where were the VT graduate students when they submitted their ETDs and completed these survey questions? Nearly two-thirds were working on campus: 28.1% were working unassisted in a computer lab while 6.6% were in the New Media Center. Almost one-third, 29.8%, submitted from their campus offices. Nearly 20% submitted their ETDs from their residences; 18.4% were in their off-campus residences and a few, six or 1.3%, submitted from their campus dormitory rooms. 15.8% submitted their ETDs from elsewhere.

What kind of computers did these ETD authors use? The vast majority used PCs: 83.7%. Mac users comprised 13.7% of the survey respondents. Nearly a dozen others, 2.4%, used Unix-based operating systems.

The final questions of the nine asked of our graduate students submitting their ETDs, are very telling ones. In the last two questions the graduate students report their plans to publish from their theses and dissertations, and if they restricted access to their works, who advised them to do so.

Virginia Tech's ETD authors reported ambitious publication plans. Over 85% reported that they planned to publish articles, proceedings, chapters, books, and other works based on their theses and dissertations. Only 13.7% reported that they did not plan to publish anything from their ETDs. Over two-thirds of our graduate students reported that they would publish articles, and 40.7% anticipated publishing in conference proceedings. Only a small number, 26 out of 445 respondents or 5.8%, plan to publish a chapter, while a few students, 31 or 7%, plan to publish a book.

While many authors and their faculty advisors are afraid that ETDs that are available on the Web will be considered "published," only 11.4% of the authors reported that publishers had advised them to restrict access to their ETDs. On the



other hand, the graduate student authors reported that their faculty are overwhelmingly advising them to restrict access to their ETDs—235 out of 271 or 86.7%. Someone other than faculty and publishers advised 82 authors (30.3%) to restrict access to their ETDs.

#### ALUMNI'S PERSPECTIVES

In addition to having information from graduate students who submitted ETDs in 1999, the Graduate School gathered data from about 50 alumni who allowed their ETDs to be Web-accessible during the previous year. While 86.3% of the 1999 ETD authors said they would publish something from their ETDs, 42.6% of the 1998 survey respondents said they had published. Remarkably, the published authors reported that they did not encounter any resistance from publishers to accepting their manuscripts because they were derived from online theses or dissertations. Hopefully the same will hold true for the 1999 ETD authors though 11.4% said they restricted access based on advice from publishers.

The 1998 authors responding to the Graduate School's survey reported overwhelmingly that they were satisfied (66.7%) or somewhat satisfied (29.2%) that their work was more widely known and appreciated because their ETDs were accessible. Over 40% had been contacted as a result of having their works on the Web, and 88.3% were satisfied or somewhat satisfied with contacts resulting from their Web-accessible ETDs. The vast majority were satisfied (27.8%) or somewhat satisfied (55.6%) that it also helped to expand their network of research colleagues. Of those who received comments about their ETDs, 83.3% reported receiving somewhat positive comments though 16.7% said they received somewhat negative comments. Small, but equal numbers were somewhat satisfied (5) and somewhat unsatisfied (5) that having their ETDs online helped them locate employment.

Among the comments they noted that they were pleased with the ability to "share one's knowledge and research." Another popular comment was about how "incredibly convenient" it was to provide copies of their theses and dissertations by forwarding the URLs of their VT ETDs. One survey respondent expressed concern about the ease of plagiarism and copyright violation. The Graduate School is conducting another alumni survey in September 1999.

## USERS' PERSPECTIVES

In addition to survey responses from ETD authors, we collected responses from about 50 individuals between April and October 1998. Voluntarily completing a "user survey" that was a link from the library's ETD homepage (<http://scholar.lib.vt.edu/theses>), they responded to 16 multiple-choice questions. Where did these respondents work or study? 65.4% were associated with universities, 13.5% with industry, 7.7% with schools, and 13.5% worked or studied elsewhere. Who were these respondents? 44.2% were researchers, 3.8% were faculty, 11.5% were teachers, and one (1.9%) was a librarian. 13.5% categorized themselves as "other."

Of our 52 respondents, nearly one-third (32%) had relatively fast Internet connections (i.e., Ethernet, T1, ISDN, or cable modem). The majority, 54%, was connecting at slower speeds, 56 KBPS (kilobytes per second) down to less than 14.4 KBPS. The connect speed was unknown to seven of our respondents.

These survey respondents used PCs, Macs, and Unix-based operating systems in similar proportions to ETD authors. 2.4% of the user survey respondents used Unix-based systems. 13.7% of the authors used Macs, but slightly fewer of the surveyed users did, 10.2%. ETD authors were 83.7% PC users and 91.8% of the surveyed users. Only about one-fourth of the ETD users responded that they were unfamiliar with PDF.

Twenty-five percent of our ETD users were from universities that do not accept ETDs, though 27.5% responded that they did not know if their universities accepted ETDs. 40% were from universities that accepted ETDs. 18% of the respondents said they submitted ETDs, while 82% (32) said they did not. 55.2% of the user survey respondents said their universities *should* accept ETDs. Only one person answered that their university should not accept ETDs, and 20.7% had no opinion on this topic.

When asked why they came to the ETD digital library homepage (<http://scholar.lib.vt.edu/theses>), the largest percentage, 62%, said they were doing research. 21% wanted to learn about ETDs, 2% were there because of their jobs. Personal interest was the motivating factor for 15%. There was an even distribution for responses to "Were ETDs easy to find?" Seven responded "very easy" while six responded "very difficult." Ten each responded easy, fair, and difficult. Four responded to our survey before they tried to find any ETDs. In the future we may



survey most of these readers again. Over 82% said that we could contact them for further information and they gave us their email addresses.

#### FACULTY MEMBERS' PERSPECTIVES

ETDs have been a topic of conversation among Virginia Tech faculty for many years and in the winter of 1996 the Degree Requirements, Standards, Criteria, and Academic Policies Committee (DRSCAP) recommended to the Commission on Graduate Studies and Policies (CGS&P) that graduate students submit theses and dissertations electronically. Unlike the information we have gathered from student and user surveys, faculty opinions have been gleaned largely from electronic mail archive (email), minutes of the university's ETD Advisory Committee meetings, and anecdotal evidence from workshops and individual encounters.

One source is the archive of minutes of the Commission on Graduate Studies and Policies and its Degree Requirements, Standards, Criteria, and Academic Policies Committee. The minutes of October 18, 1995 show a proposal from DRSCAP that recommends that the Graduate School begin requiring students to submit their theses and dissertations electronically by the fall 1996 semester. The minutes of a January 1996 Commission meeting report six concerns: (1) phase in the requirement, (2) make resources available to support the requirement, (3) include copyright statement, (4) put in place procedures for exemptions, and (5) make allowances for hard copies and (6) adaptation to new software.

On Feb. 21, 1996 CGS&P received a document on implementing ETDs prepared by DRSCAP. After modification, it was approved unanimously. One of the modifications was to begin the requirement Jan. 1, 1997. DRSCAP recommended: (1) encourage voluntary submission prior to January 1997; (2) establish university oversight committee; (3) publicize the requirement widely; (4) provide training and workshops for students, faculty, and staff; (5) avoid significant burden on departments.

At the April 17, 1996 meeting of the Commission, an amendment was made to the Feb. 21 policy stating that ETDs would not be available on the WWW until a release form granting permission to do so was signed by both the student and major professor. The oversight committee was also charged to work with societies and publishers to encourage them to adapt their policies concerning ETDs.



As a member of the Virginia Tech ETD Team, my email archives are quite robust. They show areas of faculty concerns focussed on three main topics in 1996: (1) quality of the works representing the university; (2) easy and timely access to graduate students' research; (3) impact of ETDs on future publications.

Of these concerns, only the last was reflected in the minutes of DRACAP and CGS&P. Widely available ETDs caused faculty to be more concerned about the potential impact of their students' works. They wanted their university to be represented by quality research, and, if they are going to be available, they should be timely and easy to find.

Many email messages also focused on training for students and faculty. This concern was addressed in two ways. The Graduate School has provided "ETD Workshops," informing attendees about the process of conversion from word processed documents to PDF and the online submission process. At the workshops, students frequently ask about scanning images—an indication of their planning to convert analog formats to digital formats instead of creating digital works initially.

Secondly, Faculty Development Institutes (FDI: see Endnote) introduced faculty to similar processes. In 1996, 340 Virginia Tech faculty attended week-long FDIs, subsequently receiving the hardware and software that they had been instructed in using. Seven-hundred faculty had participated in the faculty development initiative by the end of the first year of the ETD requirement.

The third issue of concern to faculty in 1996 was the impact of future publications and whether an ETD would be considered published if it was available on the Worldwide Web. Unable to definitively answer this for every publisher, Virginia Tech responded to this concern by offering restricted access options. This was to give publishers time to become familiar with ETDs and reduce the perceived threat that ETDs posed journal sales. In addition to the option of unrestricted, worldwide access, students and their advisors have the option to limit access to the university community or to withhold access from everyone, both for a limited amount of time.

In 1997 the issues of concern to Virginia Tech's faculty are documented in articulate, prolific, and heated email transmissions. Many faculty felt the ETD requirement had been handed down to them by the university administration without any opportunities for faculty input. Evidently, faculty on the CGS&P and DRSCAP had not taken information about the ETD initiative back to their constituencies. In

addition, few faculty had read the articles published in the campus newspapers. The faculty, staff, graduate student newspaper, *Spectrum*, mentioned electronic theses and dissertations almost monthly in 1996, including workshop announcements, reporting an article about the Virginia Tech ETD initiative in the *Chronicle of Higher Education*, recounting meetings of the CGS&P and DRSCAP, announcing a FIPSE grant (Fund for the Improvement of Post-Secondary Education), and more. In 1997 ETDs were mentioned bimonthly, including articles about the feature story on National Public Radio, Microsoft and Adobe software donations to the NDLT, workshop announcements, and more.

From the email archive four themes important to the faculty in 1997 stand out: potential harm to future scholars, "cultural" differences among academic disciplines, faculty feeling threatened, and the lack of clarity about intellectual property law. Only the final issue continued to be of concern to faculty in 1998 and 1999.

The emails stressing the potential harm to future scholars had three components. (1) Would new scholars be able to get their articles, for example, accepted for publication if ETDs were considered published because they were Web-accessible? (2) Would the online works have the same quality as the earlier paper versions? The subtext of this concern was: (a) Would ETD authors continue to speculate about future research if they were afraid that their ideas could easily be plagiarized? (b) Broken Web links would detract from the work, and (c) big files would be too slow for readers to download. The last component (3) was how would UMI handle ETDs in *Dissertation Abstracts*?

Getting ETDs into *Dissertation Abstracts* was an early workflow issue resolved through programmatic email notification to UMI of each approved and available VT ETD. However, fear for students whose ETDs would be fully Web-accessible lead over 62% of our faculty to advise their students to restrict access, according to authors surveyed in 1999. 20% of our students have responded to this cautionary advice by withholding all access, and 30% restrict access to Virginia Tech. This advice has also lead to *Dissertation Abstracts* listing fewer Virginia Tech ETDs. Unlimited, worldwide access is available to 50% of the ETDs approved by the Graduate School.

Rating the quality of ETDs in comparison to other formats has not yet been address and is a topic for further study. ETDs are richer in graphics and color, very



likely due to the reduced expense producing the online version, versus the considerable cost of supplying multiple color copies that would be required for the paper format. The quality of ETDs with broken Web links has not been studied. However, many feel that an out of print article is also very difficult to find and probably leaves the researcher with a greater dirth of information than do broken links and missing Web sites. The relationship of quality to download time is questionable, especially when heavy Internet traffic slows access, but it is certainly faster than traveling to campus and parking, or waiting for a dissertation to arrive through document delivery services such as interlibrary loan or UMI.

#### **LIBRARY PERSPECTIVES**

The advent of electronic theses and dissertations also raised a cadre of issues and responsibilities for libraries to address, including improving access to information, maintaining the information server, and archiving. With ETDs libraries became better-stocked digital libraries, provided more timely access to information, served more users without increased staffing, and saved shelf space. Some issues are new to libraries but most require mapping traditional services and resources to digital resources and electronic services.

The library is a traditional source of information, but it has not necessarily established itself as the locale for computer training. This is changing, not because of, but along with, the advent of ETDs. There are typically two types of library computer labs: those supporting learning about access to resources and those supporting training and resource development. Some libraries provide both, assisting authors to prepare ETDs and introducing users to new information resources such as ETDs. According to the author surveys, library instruction has been somewhat effective based on author use of the New Media Center. Faculty also receive training in the use of technology to prepare electronic works for publications and they are introduced to the library's new information resources.

Librarians with increased instructional responsibilities assist users throughout the full information cycle, from discovering to communication to submission, possibly for online access and publication. In addition, computer labs in libraries can familiarize graduate students with asynchronous collaboration tools, enabling them to work with local as well as remote resources. As graduate students become more



sophisticated software users (as are today's incoming university freshmen), libraries may provide authoring spaces with less intervention.

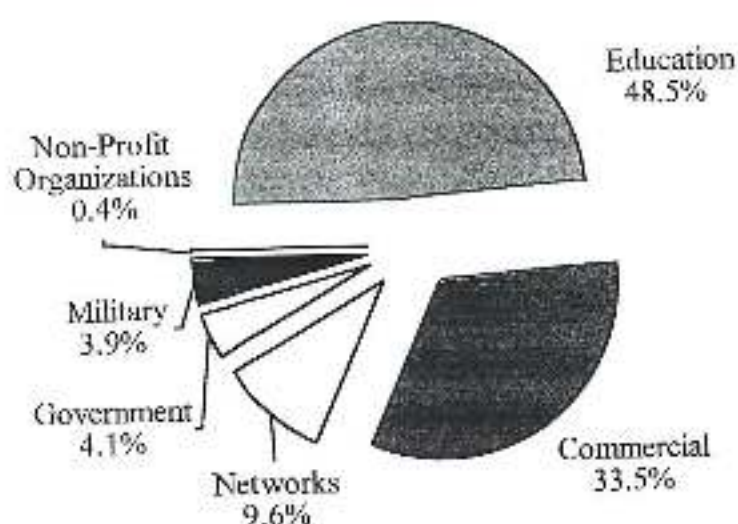
With ETDs libraries enable universities to unlock intellectual property. Unlike the research described in printed and bound theses and dissertations that has been essentially hidden because it circulates so infrequently from library shelves. Similarly, very few dissertations, available through *Dissertation Abstracts*, have been requested frequently enough to garner royalty payments. Virginia Tech's ETDs are available to a wide community of users unhampered by the hours the library is open and this results in far more frequent use than the paper or microfilm formats.

ETDs Requested			
<u>1996</u>	<u>1997</u>	<u>1998</u>	<u>Jan.-Aug. 1999</u>
4,600	72,854	244,987	476,313

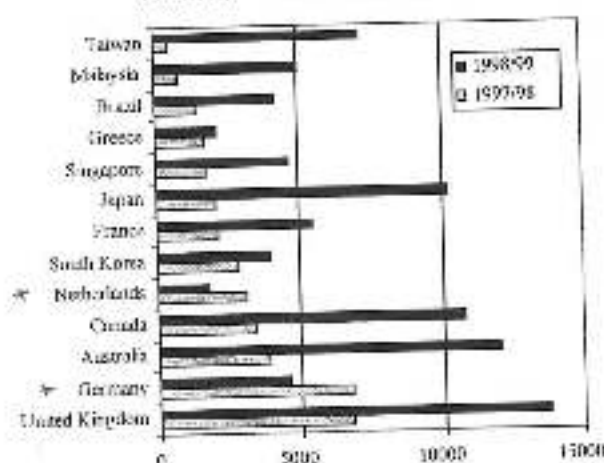
Computer log files reveal that not only does the library provide increased access to its ETDs, its clientele is broader than the local educational domain and now reaches commercial and government domains.

ETDs are labor saving devices! Increased use does not require increased library staff time because ETDs do not have to be mailed to *Dissertation Abstracts*, bound, labeled, security stripped, barcoded, checked out and checked-in, shelved and reshelved.

### With ETDs, Libraries Serve More Users in the United States



### With ETDs, Libraries Serve More International Users



Computer programs "move" submitted works through the approval process to availability and into the archive. Even without deriving the cataloging record from the online text, eliminating handling of multiple paper copies of each title could save 73.3% of the cost of processing paper dissertations.

Libraries continue to provide access to ETDs from within their buildings for patrons. At many research universities, libraries also often provide "better" downloading times through Ethernet connections. Therefore, libraries provide Internet workstations throughout their systems for students to access completed ETDs as well as supporting computer labs for authors to prepare these works.

In the future, libraries may be the principal locale for the variety of software necessary to view multimedia ETDs. Standards for software used by ETD authors are developing by default, rather than through mandates or requirements.

#### Multimedia Recommended / Used in ETDs

<u>Formats Recommended</u>	<u>Formats Used</u>
<b>IMAGE:</b> bmp, dxf, gif, jpg, tiff	.dxf, .gif, .jpeg, .pdf, .tiff
<b>MOVIE:</b> avi, mov, mpg, qt	.mpeg, .qt, .eps
<b>SOUND:</b> aiff, mcd, wav, mp2	.aif, .wav
<b>TEXT:</b> pdf, html, SGML	.pdf, ETD-ML
<b>OTHER:</b> Macromedia, SGML, XML	Authorware, Director, xls

Some library users are uncomfortable when libraries do not have copies of ETDs in paper, preferably, but also microfilm formats. There are two important factors, however, that many library users do not consider. One is that some ETDs were not designed to convey the same message as text on paper (see the list of media above). Over time users will become accustomed to this media, especially when they enjoy, for example, the benefits of online works having multiple simultaneous users. Advantages also include the end of mutilated, destroyed, or stolen materials requiring expensive replacements or elimination of the works from library collections. The security of multiple copies of every electronic work should alleviate some of the tension created by not having these works also available on the shelves.



ETDs can become accessible immediately, once the appropriate university unit completes the evaluation of the work. There do not have to be any processing delays as with paper. In addition, a library's database of ETDs can be reindexed frequently so users can reliably search and retrieve newly available works.

Quality as well as timeliness of cataloging records for theses and dissertations has improved. When the bibliographic information from the title page is available online, catalogs can provide easy access to this information easily. When the abstract is also online, more information can be included in the catalog's bibliographic record. Library systems eventually will programmatically derive the cataloging records directly from the digital works, and with more computing power, some institutions will index entire ETDs and every word in these works will be a point of discovery for Internet users.

While libraries improve workflow and take advantage of the ease of providing prompt access to ETDs, they also have the continuing responsibility to maintain long term access and to archive these works in their final form. Maintaining the server for ETDs need not be a huge added responsibility for libraries that already maintain computers for storage and access to other digital works. Libraries, especially those that have a tradition of strong internal systems support, also have a responsibility to preserve ETDs at the same time that they provide continuous customer support.

One of the best security practices is for multiple agencies to reciprocate online archiving, and the most effective way of doing this is through "mirroring." While this has as yet not been formalized, it is a concept under discussion among members of the Networked Digital Library of Theses and Dissertations (NDLTD). But well managed, copies can also effectively back-up online resources. Many academic libraries are working with their computer science departments where research and development in digital libraries holds great promise for the future. But even so, the fact that PDF is a pervasive format used heavily by the federal government as well as many commercial publishers, means that university libraries will not be alone in trying to solve migration issues.

#### CONCLUSION

Graduate students who develop ETDs, will be better prepared to enter and fully participate in academia in the Digital Age and to be more effective digital library

users even outside of the academy. They will be accustomed to creating online resources in addition to finding and using them. As novice academics they will be better prepared to submit articles electronically if they have already learned to develop and submit their digital graduate research. Libraries can be the locale where the full cycle of information takes place, from creation to access.

The speed with which research becomes available, combined with easy access on the Internet and Web, has resulted in many graduate works getting the exposure and use they deserve. The scholarly community, both here and internationally, as well as American government and commercial users, welcome timely and free access to the research available in ETDs.

Nearly 60 universities throughout the United States and internationally (see <http://scholar.lib.vt.edu/theses/NDLTD/members.html>) are participating in the evolution of electronic theses and dissertations by joining the Networked Digital Library of Theses and Dissertations. They are adapting traditional procedures and workflow to incorporate local authors' direct submissions into timely, worldwide access. The computer programs and scripts written by the Digital Library and Archives staff to process ETDs are available without charge. While many are anxious about the unanswered questions remaining (such as the durability of digital works), many more are satisfied that ETDs receive increased use instead of adding to the thousands of theses and dissertations gathering dust from lack of use in university archives.

Through the perspectives of students, faculty, readers, and libraries, this paper has reviewed some of the immediate, near term, and long term issues related to electronic theses and dissertations. At Virginia Tech we have garnered experience from over five years of direct contact with many cultures, both within and outside the academy, and they have contributed to the evolution of theses and dissertations from paper to digital formats and their dissemination throughout the world.



## DATABASE STUDIES

### Methodology

The first attempt at a study of the transportation literature reviewed what types of materials were in the appropriate databases. This would be the first resource used by a librarian for a literature review and an excellent research tool for the practicing transportation researcher. Many projects would start in this manner. Once a listing of pertinent resources was retrieved, the availability of these resources would be of vital interest. Items in a database might produce an interesting bibliography but the primary resource is essential for quality research. Therefore, not only were the types of resources indexed in the databases important, but the ability to actually obtain the resource in a timely, cost effective manner was of interest.

To determine availability of the records indexed in the database a sample of database records was collected in 1995 and another in 1996. The methodology of each study was identical. Each took a random sample of records for the same topics and years, coded the record for format, language, and country of origin for reports. Next step was to search each record in the Online Computer Library Center (OCLC) for holding information. OCLC is a library utility that houses records of resources from over 30,000 libraries from around the world. Current statistics from the company's web site state that more than 38,000,000 records in over 400 languages are in the database which also includes a listing of libraries that own an item. It is a major source for locating libraries from which to request an interlibrary loan.

The first study was based on the TRIS transportation database from the Transportation Research Board, at that time one of the major transportation databases and widely available through the DIALOG Information Services group of databases. Two hundred seventy eight records resulted from the subject searches. The second study was performed in TRANSPORT, a new database that included TRIS and the International Road Research Documentation database from the Organization for Economic Co-operation and Development (OECD) and the European Conference of Ministers of Transport's TRANSDOC database. The subject searches in TRANSPORT resulted in 540 records.



## Results

Much of the study covered what could best be described as the basic demographics of the database records. For example, the format types were tabulated for both the 1995 and 1996 studies and are summarized in Table I. English was the major language representing 91% of the first study and 83.6% of the second. Take in Table I

The Luxembourg definition which states 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" was used. (GreyNet Website) In the two studies, 73% and 85% of the citations were to grey literature.

However, it was the OCLC results that were the crux of the study. OCLC represents the largest, most commonly used resource for the sharing of resources in the U.S. and Canada and does have a broad international perspective. The most important question of the study was whether citations from the database searches were actually available in the U.S. and Canada for fast and cost effective use. OCLC provided the means to answer this question from an American perspective.

The results of both studies were similarly grim. Overall, 77% and 76.7% of the records were held by at least one U.S. or Canadian library. Initially that might appear positive but in reality roughly one in four of the records was not readily available, not an overwhelmingly encouraging number.

However, when the records are segregated by language, a more disturbing trend becomes evident. Zero holding locations for the U.S. and Canada now represent 64% and 71% of the records. Additionally, 8% and 17% are only held by one location.

Implications were clear: on the one hand, the best information available was needed to advance transportation studies into the next century and this included many resources from international resources, but on the other hand, indications were that a number of resources were not readily available. Research and observation have indicated that timely access is crucial in technology, yet many resources were not readily available to the researches.

## BIBLIOGRAPHY STUDY

### Methodology

The next step was to expand the study and go beyond the records indexed in a database to the actual literature. What were the transportation researchers citing in their published papers? Were they able to obtain the resources they needed, regardless of language and location of publication? Did their actual references mirror the databases or did they more closely reflect national holdings?

To provide a snapshot of this aspect of the transportation information, eleven papers from various publications from the Transportation Research Board (TRB) and FHWA were analyzed. The eleven included papers from the *Transportation Research Record* and the *Transportation Research Circular* and reports published by the FHWA. The papers were published during 1997 and analyzed in December 1997. A total of 367 citations were included in this study.

The results can be summarized as follows. Books and conferences represented 39.3%, journal articles 30.7%, reports 28.7% and theses 1.3%. English language publications totaled 96% of the references. In the report format, only 12.5% of the citations were to reports from outside the U.S. Grey publications totaled 86% of the references.

It was evident from these snapshots, taken from a number of vantage points, that there was indication of a potential problem with obtaining and using international transportation information with inherent implications that were highlighted in the *Value of Information* report and other reports on this topic. It also was becoming evident that the U.S. was not collecting international resources at a significant level and the researchers were not citing these resources in their publications. These were important issues that several organizations, including TRB, the U.S. Department of Transportation's Federal Highway Administration, the American Association of State Highway and Transportation Officials (AASHTO) and the Special Libraries Association's Transportation Division, were considering.

However, several other questions were raised from these studies. Were our international colleagues facing the same problems? What resources did the researchers abroad use and were these resources available to U.S. and Canadian researchers?

With the support of the FHWA and AASHTO, a study of eleven transportation centers in nine countries was conducted during the second half of 1998. These centers represent most of the major European centers. Several developing countries which have relationships to U.S. and other European transportation centers and who utilize technology transfer centers in the U.S. and Europe rather extensively were also included in the study. They could provide an indication of information needs, utilization and production of information from a different perspective.

The transportation centers visited included: Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland; Estonian National Road Administration, Tallinn, Estonia; Federal Highway Research Institute (BAST), Bergisch Gladbach, Germany; Finnish National Road Administration (FinnRa), Helsinki Finland; Swedish National Road Administration (SNRA), Borlange, Sweden; Swedish Road and Transport Research Institute (VTI), Linkoping, Sweden; Technical University of Delft (TUD), Delft, The Netherlands; TNO-Delft, Delft, The Netherlands (a for-profit agency); Transport and Road Research Institute, Kuanas, Lithuania; Transport Research Centre, Bmo, Czech Republic; Transportation Research Laboratory (TRL), Crowthorne, The United Kingdom.

A list of questions was generated and reviewed to gather appropriate demographic information. A random sample of locally produced reports was selected and the bibliographies and title pages were copied for analysis upon return to the U.S. References were collected from EPFL, BAST, FinnRa, SNRA, VTI, several Dutch federal agencies, and TRL. Reports were not available for study from the Czech Republic, Estonia or Lithuania.

## Results

During the interviews all interviewees stated that most of the center's researchers were conversant with at least one language beyond the native language. In many cases, the researchers were comfortable with several languages. English was widely read and/or spoken and English language materials were not a problem for most of the researchers. In addition, some level of expertise in many languages was available at the centers. Therefore, translation assistance was frequently in the center itself, at least to the level needed to translate needed graphs and tables. While translation services generally were used when needed, they were not used



frequently due to the in-house capabilities. This contrasts significantly with the more mono-lingual U.S. research population.

The citation analysis was based on the collection of reports collected from the centers and U.S. libraries. A total of 35 reports were collected from Finland, Germany, The Netherlands, Sweden, Switzerland and the United Kingdom yielding a total of 627 references.

Format types are summarized in Table II

Take in Table II

The composite format percentages are reports (58%), books and conferences (26%), journals (12%), theses (1%) and others (3%). The 'other' format type includes web pages, brochures and catalogs.

In most cases native language resources represented the greatest percent of the total. The Netherlands was the exception. The percent of native language to total citation numbers was Finland 78%, Germany 68%, The Netherlands 29%, Sweden 84%, Switzerland 97% and the United Kingdom 93%.

The number of languages represented in the cited references ranged from two for Germany (German and English) to seven for Finland (Danish, Dutch, English, Finnish, German, Norwegian and Swedish). Language representation as a percentage of the total was English (36%), Swedish (19%), German (19%), Finnish (14%), French (6%), Dutch (4%), Italian (2%), and Danish and Norwegian (each .4%).

The last analysis of the report references was a determination of the availability of each resource in the U.S. or Canada. Each citation was searched in OCLC and the number of U.S. and Canadian libraries that were listed as lenders was noted. Results are summarized in Table III.

Take in Table III

## DISCUSSION

The findings of this study of transportation bibliographies, published by a number of international centers, mirror the snapshot of bibliographies from TRB publications. First, centers have a very strong tendency to cite resources in their own languages. In many cases this translates to locally published reports. Although all interviews indicated that the local researchers read several languages, the native language was almost always the most often used. In this study, non-native language

citations rate was 7% for the United Kingdom, 16% for Sweden, 22% for Finland, 32% for Germany and 71% for the Netherlands. Switzerland had only 3% of the citations in non-native language, but since French, German and Italian are official languages, this figure is somewhat misleading. Still, the majority of citations were to native language publications, negating some of the benefits of superior language skills.

The second most common language is English and represented 36% of total citations. Most collections had a number of U.S. publications in their collections, most notably U.S. Department of Transportation reports and the various publications of the Transportation Research Board.

Second commonality is the preponderance of grey resources. Publications from commercial producers are included in the citation analysis. Most of these citations are to the major commercial journal publishers. However, the majority of citations fall within the definition of grey literature written at the Luxembourg conference in 1997. In the earlier database reviews, 73% of the records in the TRIS study and 85% in the TRANSPORT study were to grey literature. In the TRB/FHWA bibliographic reference analysis the number was 86%. The results from the references cited in the international reports are similar to the percentages found in the similar, with 89% of the citations to grey literature. Table IV summarizes the percent by country.

Take in Table IV

While there is some variation between the studies, most notably the TRIS study, the other three hover at 85% or more references to grey literature. This is a high number and is likely one factor in the difficulty in obtaining transportation literature. An analysis of reports by specific country was not computed, however, a cursory glance at the report literature indicates that the native reports and the U.S. reports are the most popular, indicting the prevalence of local reports and those from the U.S. and a relative dearth of non-local reports. Even those countries that cited a number of nations tended to use those from surrounding countries.

One issue that was made clear in the interviews was a different point of view concerning free access to information and the use of the World Wide Web. Both are linked since they are linked to the issue of fiscal responsibility.

Most of the transportation centers are responsible to some degree for income generation or cost recovery. The sale of their publications is one means to generate income. While they may be willing to link their publication list from their web site,

ft GL



they are not able to link the full text document for free access. Instead, reports that might have been exchanged with other libraries or sent gratis are now sold at significant cost to the buyer.

In a similar manner, the Web is not viewed with the same enthusiasm by some of the European centers. The Web sites are created for promotion of the centers and their products, not free access to the products themselves. Their publications are copyrighted and generate income for their facility. This stands in opposition to the U.S. trend of placing full text documents on local web sites or from the more comprehensive National Transportation Library (NTL) web site maintained by the U.S. Department of Transportation's Bureau of Transportation Statistics. Thousands of reports, dozens of databases and directories are freely available in a searchable web site.

Any hope that similar virtual libraries would be created by the international centers faded during the interviews. There is little interest in loading full text reports and databases for free access. Since most of the reports are copyrighted, they cannot be mounted by the NTL. Therefore, access will continue to be through traditional collection development routes.

Collection development has been stressed by increasing demands, increasing number of resources published, increasing prices and decreasing budgets and limited space. However, in transportation this is also linked to the large number of resources that are grey literature. [While grey literature is far from impossible to collect, it does present another level of complexity to the already difficult collection picture.] More items need to be searched and ordered individually and some of the resources are difficult to locate through routine procedures. This is one of the reasons the anecdotal information stated that so many resources could not be located for loan or purchase.

The general picture from the interviews and the analysis of database records and cited references was that there is a wealth of valuable resources to support research and improve transportation studies. However, that wealth of resources is closely linked to a poverty of access points for researchers to obtain the resource, whether in print or in electronic format.



## RECOMMENDATIONS

While the picture that is painted by this multi-faceted study is rather dismal, there are several hopeful signs. First, there is a great deal of interest by various organizations to formulate a solution and devise and implement an action plan. The National Cooperative Highway Research Program (NCHRP) has funded a proposal to investigate the collection, storage, dissemination, marketing and translation of non-English transportation resources. It will initially focus on French, German and Spanish language resources but could be expanded at a later date.

While this process is underway, other initiatives are needed. A large, formal strategy is recommended, there are more 'grassroots' actions that can make a significant difference in a short time.

First is the need to know what is really available in the Americas. There are numerous collections in state departments of transportation, specialized academic collections and transportation organizations that are not cataloged on the web. The simple loading of a library catalog or a collection list could provide the exact information needed. These could be searchable through the search engine on the NTL, with a link to borrowing information. There are a significant number of libraries that are not and may never be listed in OCLC and their collections are known to a few. This is a low cost solution to a basic problem. The benefit is free access to the catalog record for all with Web access.

Second, the complexity and need for a systematic collection of international resources needs to be acknowledged and addressed. There are still opportunities for exchange of materials. While there are probably few U.S. institutions that can afford the exchange costs or storage costs of a exhaustive collection, some can take on the responsibility of selective language, country or topic collection and widely share this information. There is the possibility that some of this may be funded by a coordinated national strategy recommended through the NCHRP project.

The statistics indicate the importance of grey literature to the field of transportation. Locally produced reports are heavily cited in bibliographies. The relatively ubiquitous U.S. Department of Transportation reports are well represented also. This is a strong indication of the importance of this very primary source of research information, yet there is a definite difficulty in obtaining non-U.S. reports in the U.S., and to some extent other countries. A wealth of technological innovations,

new techniques and research dead end warnings are going unheeded due to lack of access to the literature.

Increased attention must be paid to the non-commercial aspect of the transportation literature. While the major organizations and government publications are relatively well known and collected, materials of a more local nature and other publishers need to be monitored and collected. Items such as brochures and catalogs can provide valuable information, but are not always kept, and if kept, may not be cataloged.

Third, while there are realistic limitations to the use of the Web, international co-operation should be pursued. There are some items that can be made available and even catalogs and lists that are maintained are informative.

In a similar manner, the inclusion of as many resources as possible in the transportation databases should be encouraged. The wider the exposure of the resources, the greater the likelihood that someone will want the item and it will be included in a library collection and available for loan.

## CONCLUSION

Transportation literature relies heavily on report information. Since these reports often are, by the nature of the funding and utilization patterns, the product of government agencies, they are in the accepted definition of grey literature. Many people have studied and described the intricacies of grey literature. Many have emphasized the vital importance of this type of information. Few fields have such a high percentage of information in the non-commercial arena. Transportation presents an interesting, intricate and important subject to develop efficient and systematic procedures for the identification, acquisition and marketing of grey literature. Attention to this process may be critical in efficient progress in the field of transportation.

Table I Format Types

Format	1995 Study (TRIS)	1996 Study (TRANSPORT)
Journals	28.4%	22.2%
Books/Conferences	21.6%	41.7%
Reports	49.6%	35.7%
Theses	.4%	.4%

Table II Format Type by Country

	Finland	Germany	The Netherlands	Sweden	Switzerland	United Kingdom
Reports	31	33	36	88	55	64
Books/Conf.	52	19	21	10	7	27
Journal articles	10	62	18	6	10	12
Journals (unique titles)	7	27	7	6	8	10
Theses	0	0	0	1	2	3
Other	0	9	1	4	3	1

2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100

in USA or Canada

Table III Number of Holding Locations as a Percentage of Citations

	0	1	2-5	6-10	11-25	26+
Number of Citations	379	38	32	12	19	47
Percentage	72	7	6	2	4	9

Table IV Percent of Grey Literature Cited in Locally Produced Reports

Finland	Germany	The Netherlands	Sweden	Switzerland	United Kingdom
91%	80%	90%	97%	84%	91%



as diverse as health, engineering, chemical and environment. The subject of safety also constitutes a vast and diversified field of research for both research centres and institutions operating in the sector.

The importance attached to information has resulted in the production of a great quantity of documents, most of which share the typical characteristics of Grey Literature (GL). The retrieval and management of such documents is, however, critical due to their limited availability, the difficulty in pinpointing document producers and the lack of bibliographies gathering information on this type of documentation.

This paper illustrates the results of the first phase of a two-year project funded by the Italian National Institute of Occupational Safety and Health (ISPESL), the technical-scientific body of Italy's Health Ministry, which is fully involved in the promotion of new initiatives in the sphere of prevention and safety in the workplace. The main aim of the project is to develop a database to be distributed on information networks, containing Italian GL in the field of occupational health and safety.

The design of a data base entails not only the selection of the best technological solution, but above all decisions on the type of information that the database should contain, acquisition methods, the organisation and management of documentation as well as an analysis of actual and potential users. Due to the inherent characteristics of GL, the data acquisition phase is critical: producers must be identified along with their methods of publication and distribution, even if these are limited and informal [5, 6].

The first phase of the project therefore consisted of a questionnaire-based survey aimed at potential GL producers, in order to identify sources, types of documents produced, publication, dissemination and bibliographical treatment.

The survey also has the secondary aim of finding out whether special types of GL documents are produced in the health and safety sphere which contain information specific to this sector.

This paper is divided into three parts. The first part describes the survey's methodology, the second analyses the main results from the various sections of the questionnaire and the final part examines in detail and dissemination modalities.

## 2. Methodology

GL production generally reflects the activity of its producer, and indeed a distinction is often made [7, 8, 9] between GL produced by research centres or by government agencies or the private sector. GL is moreover produced both to *activate* communication within the same institution and to communicate with other organisations [10]. These distinctions reflect the production of special GL document types which indicate organisations covering specific roles and document types which have their own contents and objectives. This reveals the close relationship between GL production and its producer, the identification of which is the first, but essential step on the way to identify, gathering and managing these types of documents.

The difficulty in identifying the GL producer is compounded by inherent difficulties in the occupational health and safety sector in general, i.e. the number and diversity of subjects involved, the tasks and duties performed by each subject and the application of different disciplinary field. The conceptual scheme (Fig. 1) illustrates the various elements to be

Said of each  
disciplinary field

considered in the analysis of this field and their main relationships. This scheme also seeks to highlight the different points of view from which the sector can be analysed, and has been used to select the survey targets.

This sector is extremely wide, due to the fact that health and safety involves the entire production system (public administration, firms producing goods and services, etc.), in the public and private sectors (central and local administrations, private firms, etc.), whatever their size and whatever the economic sector (manufacturing, farming, etc.). Health and safety activity also involves all subjects in the production system which have to conform to the law as well as institutions with monitoring roles. Of particular importance among the latter group are those organisations and/or services set up when the current law came into force.

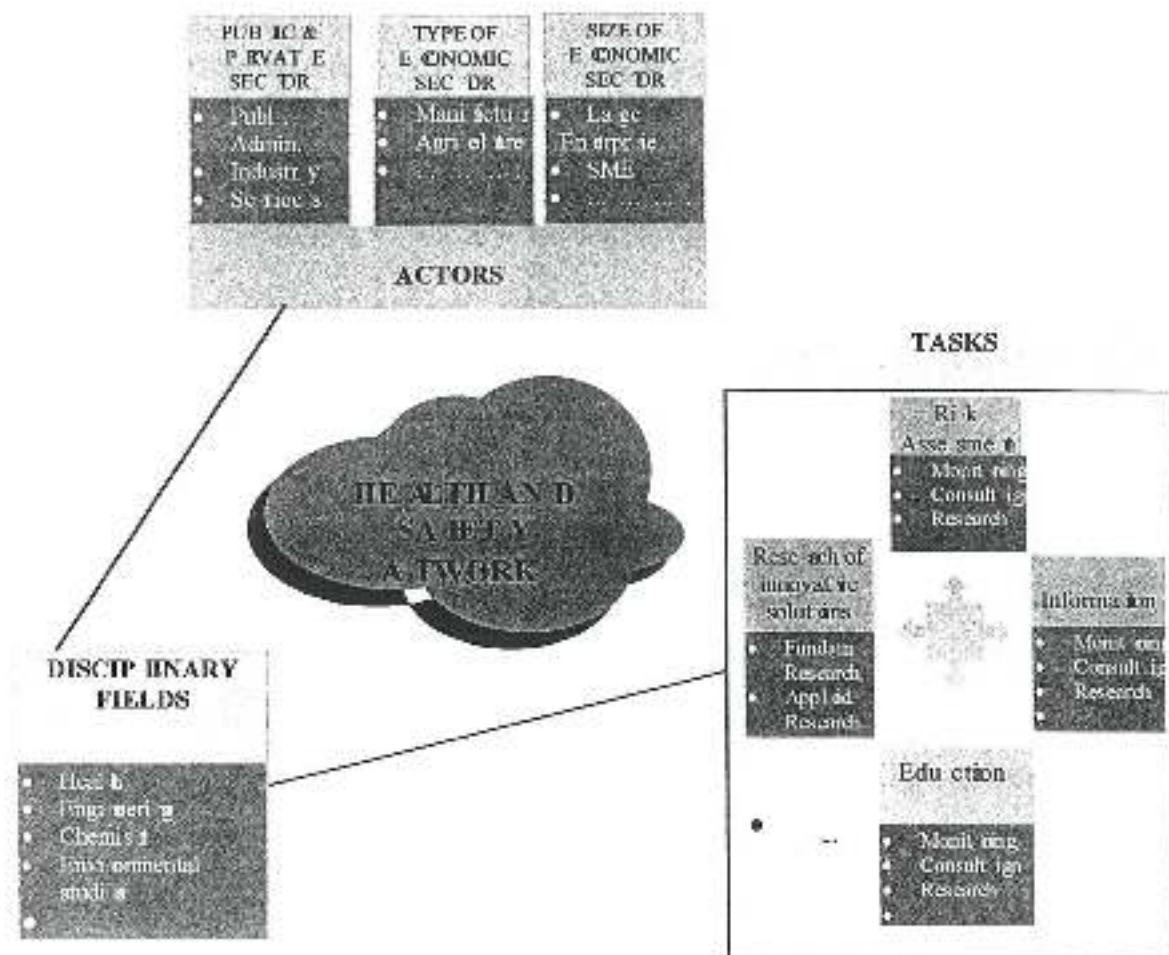


Fig. 1 - The main elements interacting in the field of health and safety

The survey on GL producers gave priority to the public sector, while the variables of the productive sector and that of organisation size were not taken into account.

The health and safety sector is multidisciplinary, ranging from the health to environmental issues. Generally speaking, however, since the law sets out to protect the health and safety of workers in any working environment, the most commonly used disciplinary competencies are medical, to safeguard workers' health, and engineering, for the correct use of equipment. This is why those organisations operating in the sphere of prevention in the medical and engineering fields were selected.



In accordance with the Italian legislation, the main functions consist of: risk assessment, research of innovative solutions, training and information. This entails research, monitoring and consulting activities, performed by all subjects in the production system, but for different ends. Organisations and public services set up when the new law came into force, for instance, perform such tasks with a view to providing guidance and ensuring that laws are enforced, while other organisations, in addition to conforming internally to laws in force, may have the aim of ensuring workers' rights (e.g. Trade Unions) or of choosing the prevention sector as a field of scientific study (e.g. Universities and Research centres).

Summing up, the area at which the survey was aimed was selected on the basis of the institutional roles of organisations and/or services and the main research, monitoring and consulting activities performed by them. Priority was also given to the public sector and to the medical and engineering areas. The main criterion adopted was that of the representative nature of the sector and that of uniformity in relation to the type of organisation to be chosen for the survey.

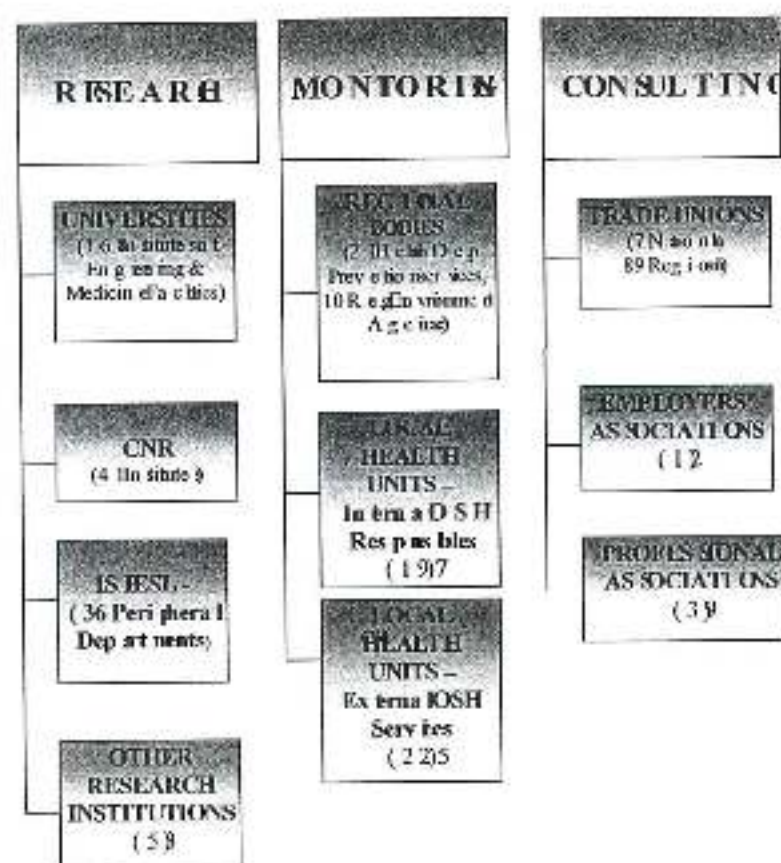


Fig. 2 - Surveyed organisations broken down by function

Of those organisations chiefly dedicated to research (Fig. 2) we have selected, from the faculties of medicine and engineering, the University institutes most closely involved with health and safety (Industrial Medicine Institutes, Hygiene and Preventive Medicine Institutes, Forensic Medicine and Insurance Institutes, Departments of Building, Construction Sciences, Material Sciences, Mechanical Engineering, Chemical Engineering, Electrical Engineering). Similar criteria were adopted when selecting CNR (National Research Council) Institutes. With reference to ISPEI, the questionnaire was sent to all of its peripheral Departments, which operate closely with local areas. Finally, the grouping "other Research Institutions" included both public and private institutions.



Organisations which carry out monitoring activities include regional authorities (Health Departments Prevention Services and Regional Environment Agencies) and Local Health Units (LHUs). Within the LHUs we have selected both the Internal Occupational and Health (OSH) Responsibles and the External OSH Services. The first are charged with the task of ensuring that the law is applied within their own institutions, while external OSH Services provide guidance and ensure that laws are complied with at a local level.

Organisations which carry out consulting activities include Trade unions and Employers' Associations. Since they represent workers and employers, they often promote initiatives to disseminate information about occupational health and safety. This grouping also includes the associations of special professional categories expressly dedicated to prevention (National Society of Prevention Workers, Italian Inter-Association Prevention Council, etc.).

To identify GL producers in such a heterogeneous environment and define their profiles, it was decided to conduct an exploratory survey using a structured questionnaire made up of 21 questions, requiring both multiple answers, in some cases with several variables, and single answers. The questionnaire was divided into five sections: production of GL documents; dissemination; bibliographic treatment; production of software tools and reception of documents produced by other organisations. In the questionnaire respondents' personal data was also collected. In this way it was possible to classify the organisations according to the variables identified in the conceptual scheme.

### 3. Main results of the survey

Data obtained from questionnaires regarding GL producers is analysed in this paragraph, as are the most significant replies from the five sections of the questionnaire (dissemination methods, management of catalogues, production of information tools and frequency with which documents produced by other Organisations are received). Refer to the next paragraph for a more detailed analysis of dissemination modalities. Questionnaire results are analysed with reference to two criteria, reflecting the aims of the survey as a whole: a) to gather information in order to draw up an organisational model from which the future database will be developed, and b) to verify whether there already exists a flow of information between GL producers concerning the exchange of documents.

54.6% of the 898 questionnaires sent out were duly compiled. Most responses came from organisations that perform research activities (66.4%). Among these, the highest response percentages were for peripheral ISPESI Departments (80.6%) and other Research Bodies (75.5%). Significant responses were received from LHUs (50.3% for Internal OSH Responsibles and 57.3% for External OSH Services), which in the past have not been studied in depth and have generally been *unaware* producers of GL, as demonstrated by the numerous contacts made when administering the questionnaire.

Table 1 - Comparison between the number of questionnaires sent, % replies received, % GL producers and % distribution of GL producers by Organisation and function

	No. sent	Responses (%)	GL Producers (%)	Distribution of GL producers (%)
<b>Research</b>				
Universities	168	60.7	56.9	17.0
CNR	41	65.9	48.1	3.8
ISPESI	36	80.6	65.5	5.6
Other Research Bodies	53	75.5	55.0	6.4
<i>Total Research</i>	298	66.4	56.6	32.8
<b>Monitoring</b>				
Regional bodies	31	32.3	70.0	2.0
Internal OSH Responsibles	197	50.3	82.8	24.0
External OSH Services	225	57.3	73.6	27.7
<i>Total Monitoring</i>	454	52.4	77.3	53.7
<b>Consulting</b>				
Trade Unions/Employer's Ass.	108	36.1	82.1	9.4
Professional Association	39	41.0	87.5	4.1
<i>Total Consulting</i>	147	37.4	83.6	13.5
<b>TOTAL</b>	<b>898</b>	<b>54.6</b>	<b>69.7</b>	<b>100.0</b>

69.7% of organisations that replied to the questionnaire were GL producers. This confirms the presence of a large number of GL producers in the field of health and safety. The highest producer percentages were for consulting organisations (83.6%) and monitoring organisations (77.3%), which also constitutes the largest single group in our study (53.7%). It is interesting to note that categories giving the fewest questionnaire replies contain the highest percentage of GL producers (32.3% for responses from Regional Bodies, 70% of which produce GL; 36.1% for responses from Trade unions and Employers' Association, 82.1% of which produce GL). It may be supposed that in some organisations the propensity to reply is determined by the very fact of being GL producers. In other cases a number of factors may determine the propensity to participate in surveys, for example being accustomed to compiling questionnaires or the ease in identifying the person best suited to supplying the requested information.

### 3.1 Types of GL Documents

An analysis of GL produced by ISPESI led us to include additional document types in the questionnaire alongside *traditional* GL, e.g.: Guidelines, Safety procedures and Technical evaluations. The inclusion of additional GL types derives firstly from the need to use the same terms adopted by producers, who in this case are not particularly familiar with problems associated with GL use. Indeed, whenever specific sectors are analysed [11], further types are often added to the already long list of GL document types.

Figure 3 shows the distribution of GL producers by document type and production frequency. Firstly it shows that production involves types of document closely related to the field. Most organisations produce Course hand-outs (57%), Technical evaluations (51.1%) and Technical reports (50.8%), while Bibliographies (15.2%), Newsletters (12.3%), Journals (10.8%) and



Preprints (5.3%) are produced by fewer organisations. Data regarding Preprints indicates that a very small percentage of organisations produce with a view to possible future publication in a conventional form. This confirms, albeit indirectly, that in the health and safety sector GL is the most common source of information. An intermediate value was obtained for the production of Theses (20.8%).

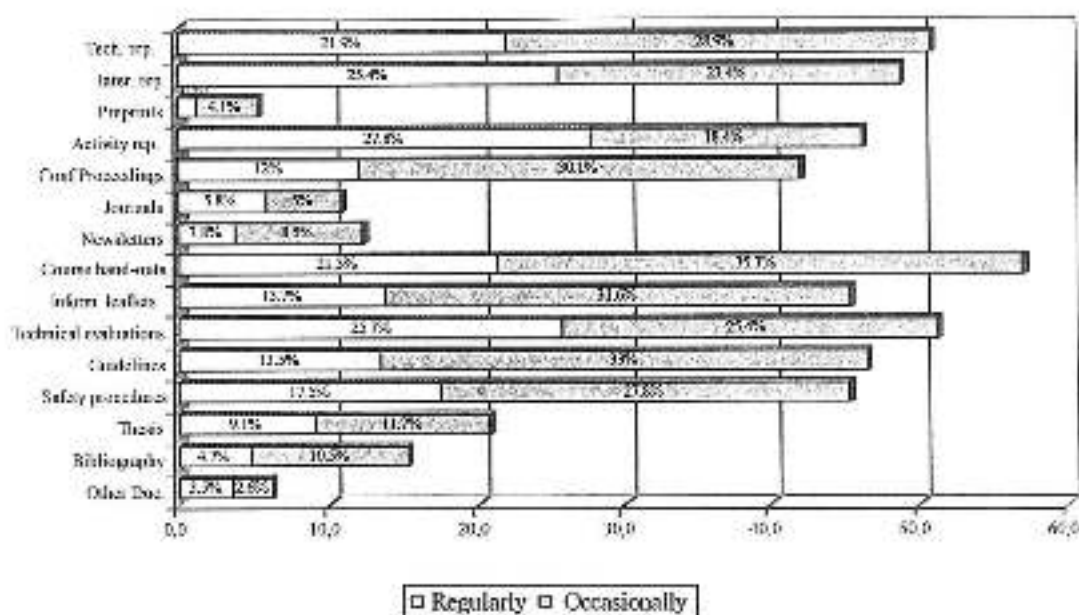


Fig. 3 - Distribution of GL producers by document type and production frequency (%)

Data on the frequency of production also show that a significant percentage of organisations produce some types occasionally, e.g. Conference proceedings (30.1%), Guidelines (33%), Safety procedures (27.8%). This data thus points to specific document types and provides indications on the various activities that play leading roles in this sector: training in the form of Course hand-outs and risk assessment in the form of Technical evaluations, Technical reports, Safety procedures and Guidelines.

### 3.2 Dissemination of GL

One section of the questionnaire had the specific aim of determining whether organisations were willing to distribute information and identifying the ways in which this is done. (see paragraph 4 for a more detailed look at this section).

The first question asked respondents whether the GL documents produced were distributed mainly to internal users or to users outside the organisation. It emerged that only 7% of respondents do not distribute their own documents, and the breakdown by organisation (Fig. 4) shows that the higher percentages are for University institutes (20.7%) and Regional Bodies (14.3%).

Data on dissemination methods confirm the subdivision by function assumed in the conceptual scheme shown in figure 1. Among the organisations performing consulting activities, 64.3% of Professional associations and 59.4% of Trade unions and Employers' Associations distribute GL documents either to external users only or to both internal and external users. As regards monitoring organisations, there was a further breakdown in distribution tendencies owing to



the specific tasks of monitoring internal safety or the safety of external structures: 57.1% of Regional Bodies and 53.7% of External OSH Services distribute GL both internally and externally, while 81.7% of Internal OSH Responsibles usually distribute GL documents within their own structure in compliance with norms obliging them to disseminate information to workers in the workplace. Among the organisations performing research activities, the role played by ISPRESL as a specialised Institute in the sector is confirmed by its internal and

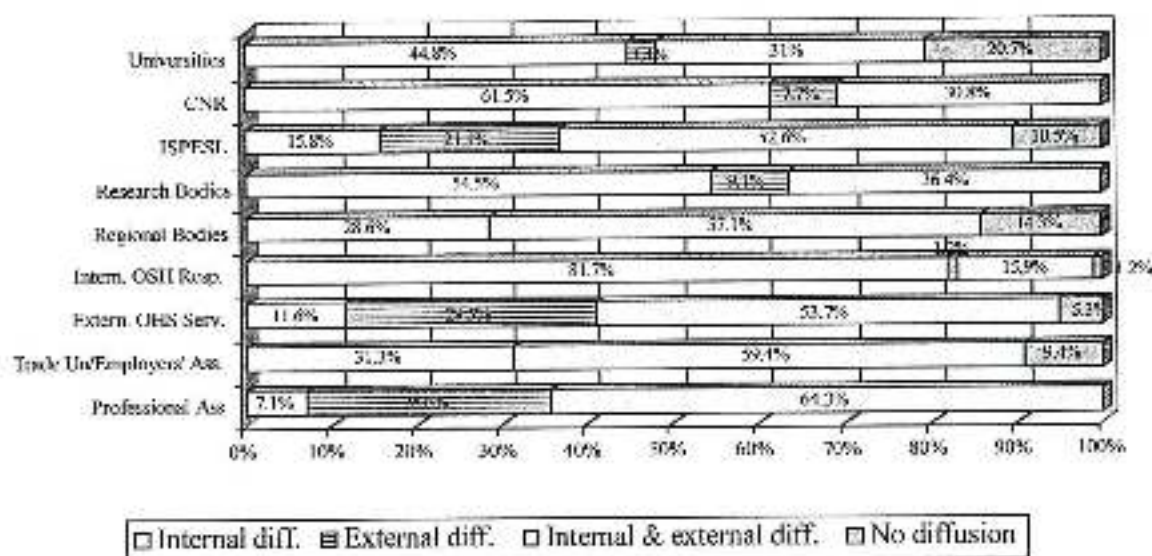


Fig. 4 - % of producers distributing GL

external distribution activities (52.6%), while for University and CNR institutes, and other Research Bodies dissemination is chiefly internal. We assume that organisations which carry out research may prefer other dissemination channels, but this should be checked through a specific analysis of conventional literature published in the field.

### 3.3 Management of GL catalogues

The respondents were asked to indicate whether the GL they produced was catalogued, and if so, which bibliographic elements were taken into consideration for document treatment. This section of the questionnaire also asked whether catalogues, databases or lists of publications bringing together GL documents had been created. This information, in addition to indicating a certain attention to the management of documentation, gave important indicators regarding the organisation of the future database. If indeed the management of GL were stable, it would be easier to acquire documentation and draw up a shared cataloguing system.

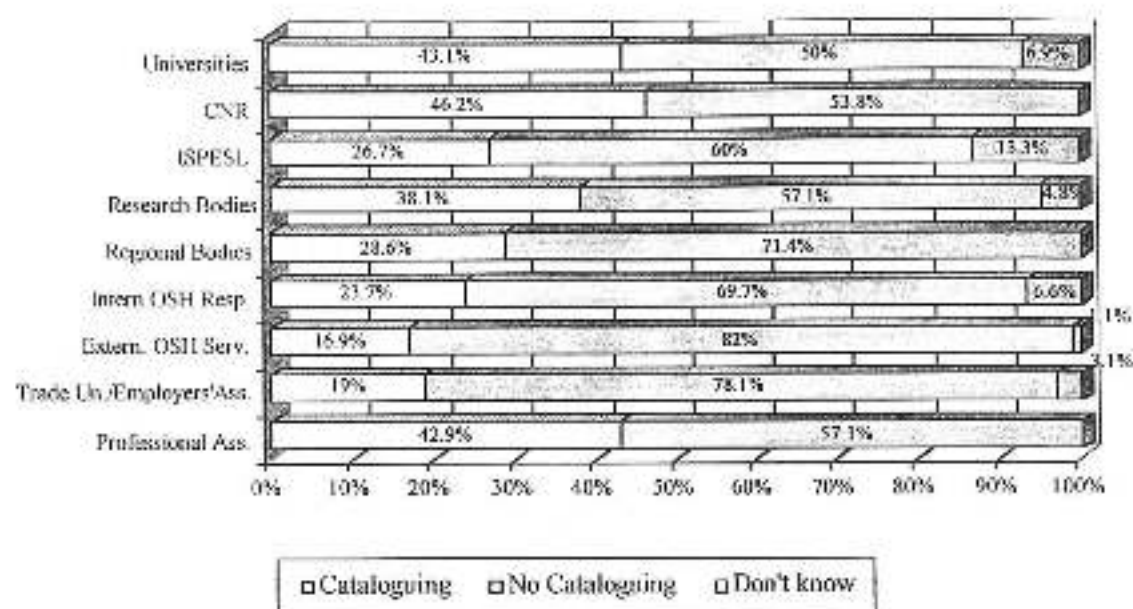


Fig. 5 Management of catalogues broken down by organisation (%)

Only 28.2% of producers actually catalogue GL. Distribution by organisation (Fig. 5) highlights some differences. The highest percentage for organisations cataloguing GL was for University institutes (43.1%), CNR (46.2%), Professional associations (42.9%) and Research Bodies (38.1%). The lowest percentages were for External OSH Services (16.9%) and Trade unions and Employers' Associations (19%).

### 3.4 Production of "grey" information tools

The fourth section of the questionnaire aimed to verify the existence of information set out in databases, or as hypermedia documents or simulation systems. In this section the respondents were also asked to indicate the media used and topics dealt with. These information tools can also be considered as *grey documentation* or, as it is often called, *grey information* [12] since it is not usually circulated outside the organisation that produces it. These tools are a move forward from *traditional* GL and, if they are produced in the health and safety sector, could be acquired or at least reported in ISPESL's future database.



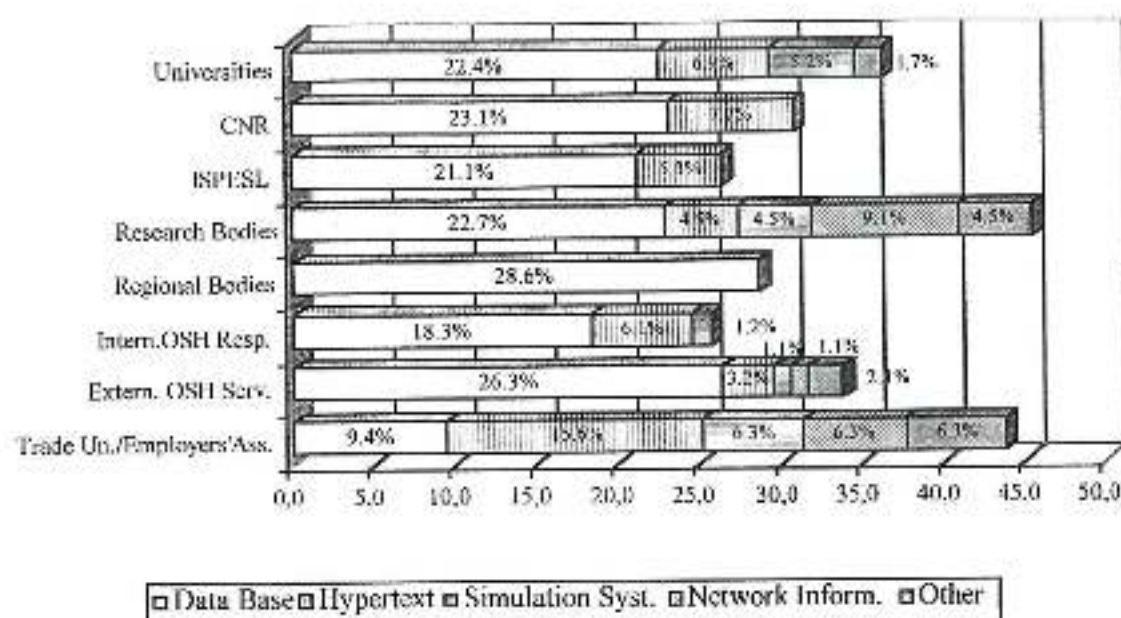


Fig. 6 - Production of information tools broken down by organisation (%)

It emerged that 101 of the 342 GI producers (29.5%) had produced "grey information tools" over the past three years. Most of these organisations produce databases, although percentages were rather low (Fig. 6). The highest values were for Regional Bodies (28.6%) and External OSH Services (26.3%). Trade unions and Employers' Associations (15.6%) specialised more in the production of hypertexts. It should also be noted that most producers of these tools use them to gather statistics and legislative information and that magnetic storage media was usually used. The chart does not show Professional associations since none of these produces any type of information tool.

If these results are compared with data on the format used for produced documents, (58.7% of the respondents produces exclusively paper documentation) with the percentage of organisations that already have a Web page (6%) or that intend to have one (30.4%), we can say that GI producers in the sector do not make full use of the potential of new technologies.

### 3.5 Reception of GI produced by other Institutions

The final section of the questionnaire concerned the existence of the final link in the information chain, namely the frequency with which producer organisations receive GI documents from other institutions. GI producers are in fact often users of this kind of document, and if they are included in an information flow, they may benefit from an exchange of documents with other organisations.



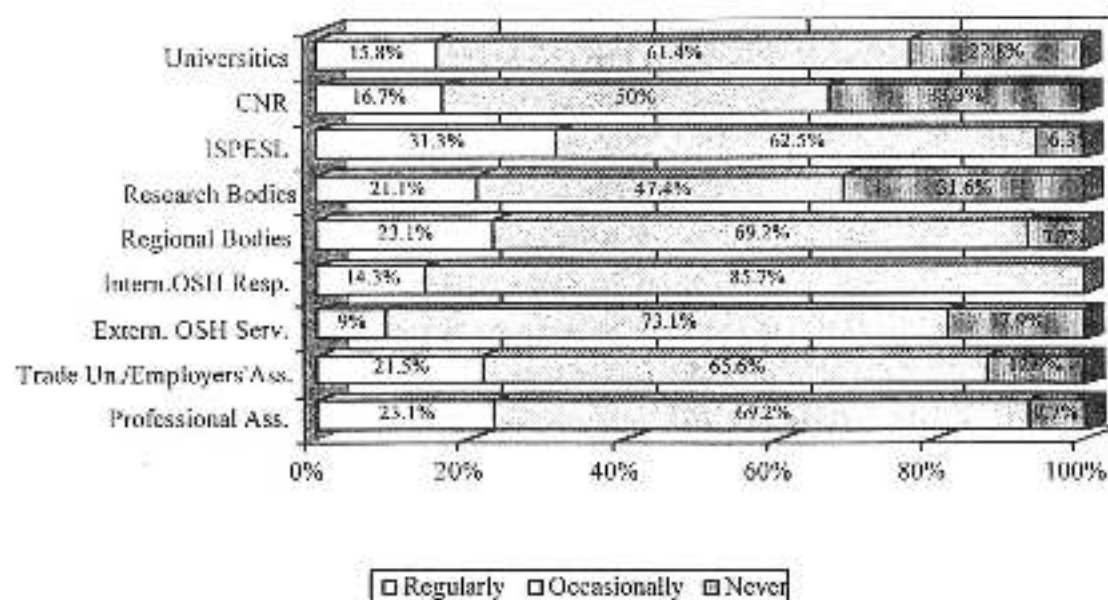


Fig. 7 - GL reception frequency broken down by organisation (%)

Most organisations receive GL from other institutions only occasionally (Fig. 7), while organisations regularly receiving GL are Trade unions and Employers' associations (32.3%) and ISPESL Departments (31.3%). Most organisations performing research activities reported that they had never received GL from other institutions (33.3% of CNR Institutes, 31.6% of Research Bodies and 22.8% of University institutes).

It should be noted that for the final question asking to indicate the names of institutions from which documents are received, many GL producers indicated ISPESL. This bears out ISPESL's role in the health and safety sector and creates the conditions for the success of the future database.

Summing up, for a large number of GL producers in the field of health and safety the data analysed up to this point reveals various profiles, sometimes relating to the organisation's institutional role and sometimes to the more general characteristics of GL. Data on GL production confirms that special types of documents are produced having sector-related information contents. Data on the dissemination of documents produced reflects, for example, the functions performed by these organisations, leading to a greater or lesser propensity to disseminate information internally or externally. Data on the non-management of documentation produced confirm the characteristics common to all GL, for which the absence of bibliographic controls has always been a major obstacle for retrieval and dissemination activities. Data on grey information tools, too, and the predominantly paper-based production of documents are proof of the partial use made of new technologies on the part of GL producers in the sector. These results will influence the organisational model of the future database.

#### 4. Dissemination of documents

The section of the questionnaire concerning the dissemination of documents touches upon one of the crucial aspects of GL. The difficulties encountered in obtaining non-conventional literature is indeed due to the non- or limited activation of communication channels and to the under-estimating of potential users. This may be caused by a number of factors, some of which are examined in the questionnaire, such as the inadequacy or absence of internal services to

support this type of activity (library, documentation centre, editorial office or staff assigned to this task), a lack of awareness about the value of information in general and about the content of GL documents in particular, as well as a lack of familiarity in using external information sources and exchanging information, albeit informally.

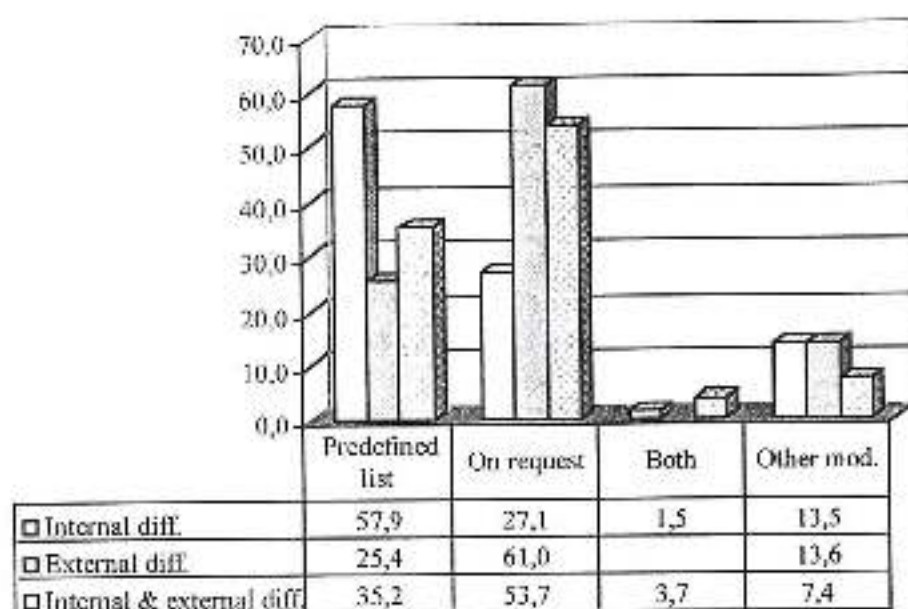


Fig. 8 - GL dissemination modalities (%)

The respondents were asked to indicate whether they distributed their documents through a predefined list or on request, or using both modalities (Fig. 8). This data is compared with results on internal and external diffusion (cfr. par. 3.2).

The use of a predefined list assumes that producers have identified a group of *habitual users*. The data gathered shows that such a list is used by most organisations (57.9%), especially for internal distribution, which, as already mentioned, is of particular significance for the dissemination of information on health and safety in the workplace. The exclusively external distribution, which represent the "expert to expert" dissemination, occurs for the 61% of organisations on request. This means that there are no fixed means of dissemination, nor are there established channels for the exchange of documents.

#### 4.1 Information flows

GL producers were asked to indicate the institutions to which their documents are distributed (Central State Administration, Regional and Local Administrations, Universities, Research centres, Trade unions, Professional associations, Private companies, own members, own employees). It was thus possible to verify whether there is already a flow of information among those GL producers chosen for the survey, and whether such a flow gives priority to particular types of institutions.

Table 2 shows an input-output grid relating organisations that send out their documents



(output documents) to those receiving them (input documents).

The institutions indicated by producers as the receivers of their documents were grouped according to their main role using criteria similar to those adopted for GL producers. Processed data did not therefore take into account replies indicating central administrations and private companies, which were not included in the survey, while organisations indicating that they distributed GL to their own members and/or employees were classed as belonging to internal flows, that is to information flows directed towards the same type of organisation.

Table 2 - Information flow among GL producers broken down by function (%)

Sending Institutions	Institutions receiving GL		
	Research	Monitoring	Consulting
Research	56.1	19.3	24.6
Monitoring	4.8	35.7	59.5
Consulting	5.7	10.5	83.8

The information flow given in table 2 shows that documents are exchanged chiefly within organisations covering similar roles, and this is particularly evident for organisations performing consulting activities (83.8%). Exceptions were organisations performing monitoring activities, which are those showing the greatest flows of information (59.5%) towards organisations performing consulting activities. This may be explained by the need to communicate the results of their monitoring and supervisory activities to consulting organisations. Organisations performing research activity receive little input from organisations with different roles, while their output is higher (24.6% and 19.3%). It should be noted however that these figures are affected by answers given by peripheral ISPESL Departments that operate closely with local contexts and, as they belong to a national organisation expressly dedicated to the sector, perform liaison activities among the various health and safety services.

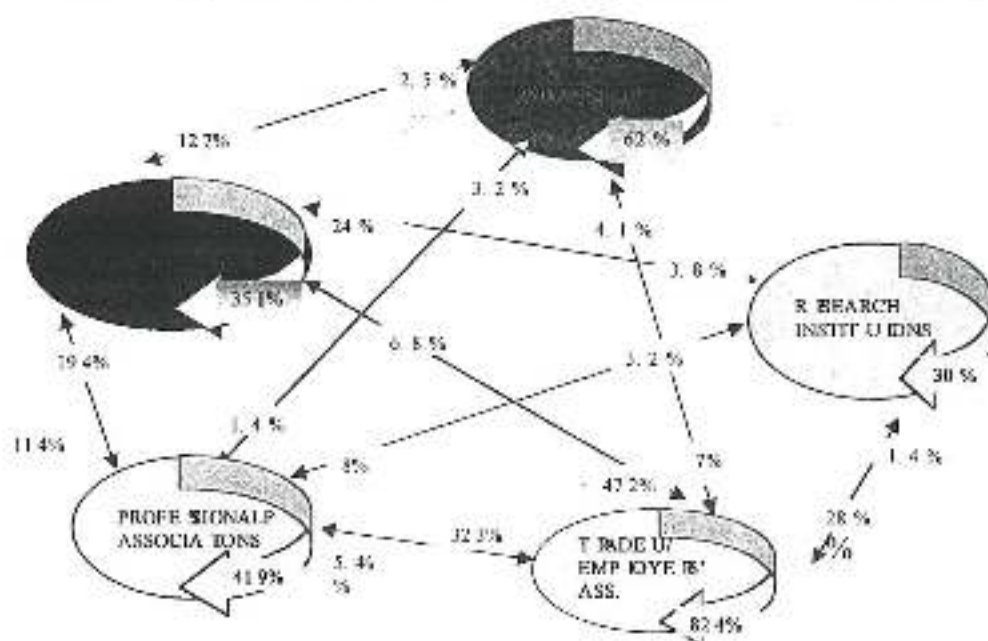


Fig. 9 - Information flow of GL producers broken down by organisation



The information flows broken down by organisation (Fig. 9) reinforces the idea of an exchange of documents limited to organisations of the same type. This is especially evident for Universities (62%) and Trade unions and Employers' Associations (82.4%). The State's Peripheral Administrations, including Regional Bodies, as well as Internal OSH Responsibilities and External OSH Services, communicate not only internally (35.1%) but above all with Trade unions and Employers' Association (47.2%).

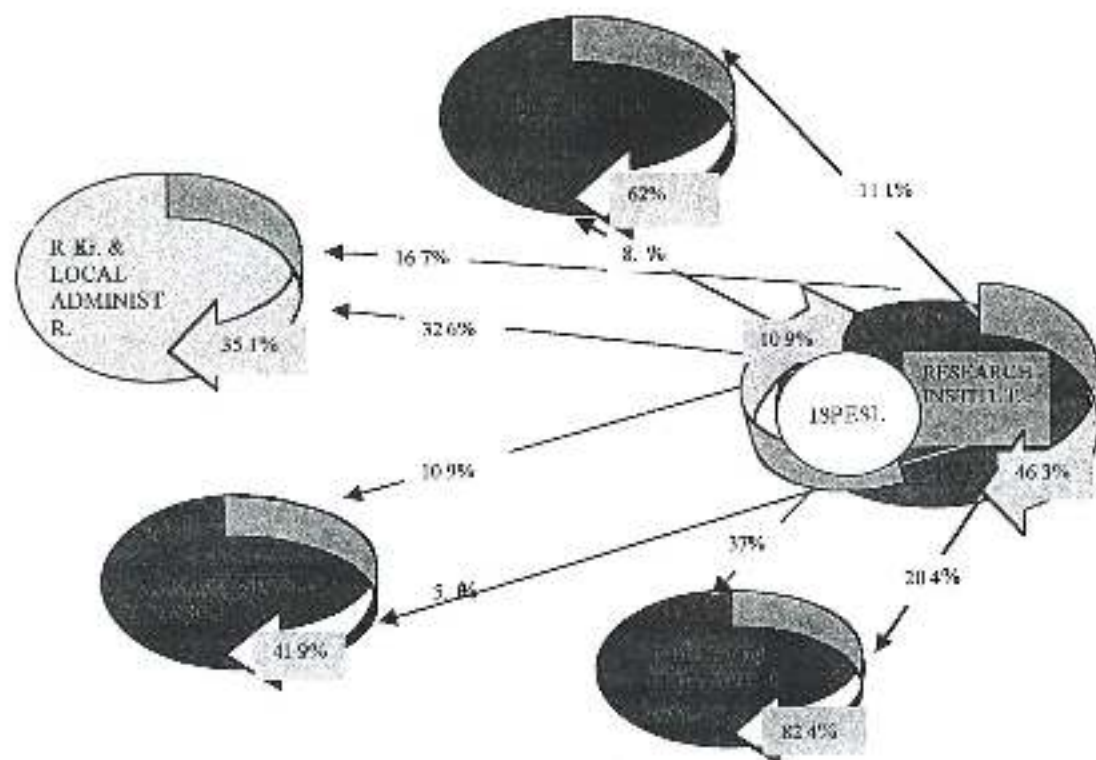


Fig. 10 - ISPESL document output compared with other organisations

Figures on Research Institutes, comprising CNR Institutes, other Research Bodies and ISPESL Departments, are again influenced by answers given by ISPESL. If one analyses individual data (Fig. 10), it appears that the information flows of ISPESL Departments are chiefly directed towards the Peripheral Administration (32.6%) and towards Trade unions and Employers' Associations (37%), while of much lesser importance are flows to Universities (8.7%) and other Research Institutes (10.9%). On the other hand the information flows between Research Institutes not comprising ISPESL Departments again highlight the fact that the exchange of information occurs chiefly among organisations of the same type: flows between Research Institutes were 16.3% higher in this case, going from 30% to 46.3% (cfr. Fig. 9). Excepting the exchange of information between Research Institutes and Universities, flows to other organisations were considerably lower.

#### *Support structures for document editing, cataloguing and dissemination activities*

GL producers were asked whether their organisations possessed structures set up to perform GL document editing, cataloguing and dissemination activities, and if so, to indicate one of those listed (library, documentation centre, archive department, editorial office, press office, publications office, public relations office). 57% of Organisations replied positively to the first question, but in the follow-up question added an unforeseen item: "document producer", turning around the meaning of the question.

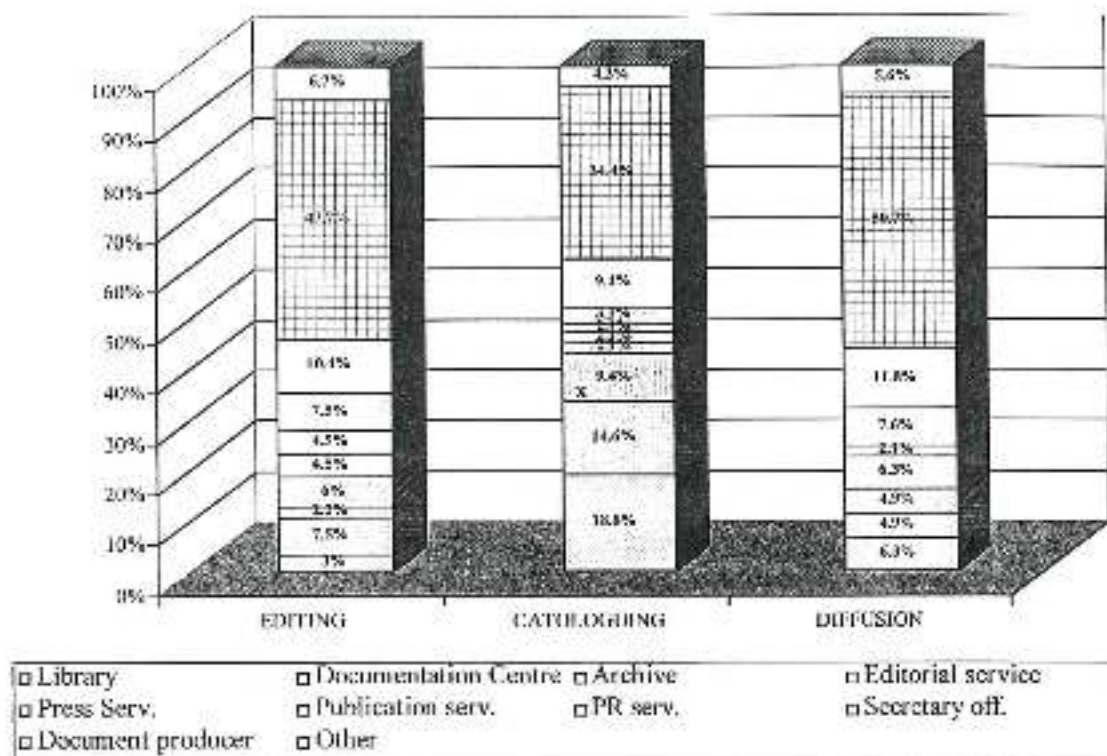


Fig. 11 Presence of departments performing GL editing, cataloguing and dissemination activities (%)

In most organisations it is the author him/herself (the "document producer") who drafts the document (47.7%) and distributes it (50.7%). Cataloguing activity is performed as follows: 34.8% by the document producer, 18.8% by the library, 14.6% by the documentation centre. This fact will render the acquisition of documents for the future database more complex.

On the whole, data regarding dissemination highlights the existence of well-identified users within organisations, receiving information on prevention and safety measures. On the other hand, external dissemination is prompted more by users requesting information than by established dissemination practices. Despite this fact, most producers indicated that they send their GL documents to other institutions, and resulting information flows are directed largely towards organisations of the same type. Only ISPESI Departments and Regional and Peripheral Administrations have opened up information channels to organisations with different roles. Finally, very few organisations make use of support services for editing, cataloguing and dissemination activities, and most of these activities are performed directly by the document's author. This result is perhaps not so surprising for GL, in which the author plays an active role in each phase of the document's life cycle, becoming the main distributor of produced documents, especially in sectors making use of online communications.

## 5. Conclusions

The survey confirms the hypothesis that a large number of organisations in the field of occupational health and safety produce GL, and that the development of a database containing such documents may be a useful tool for improving communication among different organisations, increasing information dissemination modalities and accordingly enhancing



information channels between organisations with different roles.

The survey looked at a range of different contexts in terms of the institutional mission, organisational set-up and role played by each organisation. The resulting profile of the GL producer shows specific characteristics especially for the production of field-related document types. Features shared by most organisations include the modalities used to manage and disseminate GL production. Only a small number of organisations have arranged for the bibliographic management of GL production, and dissemination is generally directed towards internal users. External distribution appears to be unplanned, information flows being directed towards organisations performing similar role. Finally, a limited number of organisations have services supporting document editing, cataloguing and dissemination activities.

These results confirm that once again the difficulty in retrieving GL and its limited dissemination derive from the low priority which producer organisations give to the exchange of information, and of the degree of awareness of its information value.

This fact has a direct influence on the design and development of the future database, since it will be necessary to choose an organisational model for the acquisition and bibliographic management of documents that can fit in with the needs of both producers and database managers. As indeed the success of the project depends on the degree of cooperation among producer organisations, which will moreover constitute a sizeable portion of future database users, a modular, flexible system will have to be developed. In particular it will be necessary to provide simple document-cataloguing methods, whenever possible, while access to the database should not be based solely on network distribution as most organisations do not yet possess such computer-based instruments.

The project is now in the data modelling and software selection phase. There are also plans to illustrate the results of the survey to those Organisations that replied to the questionnaire to raise awareness and involvement in order to build an effective communication and information network.



## References

- [1] Decreto Legislativo 19/9/94, N. 626. Attuazione delle direttive 89/391/CEE, 89/654/CEE, 89/655/CEE, 89/656/CEE, 90/269/CEE, 90/270/CEE, 90/394/CEE e 90/679/CEE riguardanti il miglioramento della sicurezza e della salute dei lavoratori sul luogo di lavoro. *Gazzetta Ufficiale della Repubblica Italiana*, n. 141, 12 Nov. 1994.
- [2] Council Directive of 30 Nov. 1989 concerning the minimum safety and health requirements for the workplace [CEC]. *Official Journal of the European Communities*, 30 Dec. 1989, vol. 32, no. L393, p.1-12.
- [3] Castriotta, M. La documentazione di tipo sanitario a misura del paziente e del lavoratore. In: *Atti del 5. Convegno Nazionale AIDA "Documentazione: professione trasversale"* a cura di C. Basili, Roma, 1998, p. 163-170.
- [4] Holzer, C.L., Hirsch, R.B. Managing workplace exposure information. *American Industrial Hygiene Association Journal*, 54 (1), 1993, p.15-21.
- [5] Auger, C.P. Information sources in grey literature. 3rd ed., London: Bowker-Saur, 1993.
- [6] Farace, D.J. Rise of the Phoenix: A review of new forms and exploitations of Grey Literature. *Publishing Research Quarterly*, 13 (2), 1997.
- [7] Hart, G.W. Aerospace and other grey literature from Nato. In: First International Conference on Grey Literature, Amsterdam, 13-15 dec. 1993. Amsterdam, Organising Bureau, 1994, pp. 18-32.
- [8] Rozkuszka, D.W. The art and acquisition of foreign official publications. In: *Official publications of Western Europe*, vol. 1 Edited by Eve Hobansson. London: Mansell Publ.Lim., 1984.
- [9] Sprehe, T.J. Government information: from inaccessibility to your desktop and back again. *Journal of the American Society for Information Science*, 50 (4) 1999, p. 340-345.
- [10] Kreitz, P.A., Addis, L., Galic, H. Johnson, J. The virtual library in action: collaborative international control of High-Energy Physics pre-prints. In: 2nd. Conference on Grey Literature, Washington, 2-3 Nov. 1995. Amsterdam, Organising Bureau, 1996, p. 33-41.
- [11] D.J. Farace. Glossary, a vocabulary of terms in GL. Grey Literature Network Service. Amsterdam: GreyNet, 1999, 60 p.
- [12] Carroll, B.C., Cotter, G.A. A new generation of Grey Literature: the impact of advanced information technologies. In: First International Conference on Grey Literature, Amsterdam, 13-15 dec. 1993. Amsterdam, Organising Bureau, 1994, pp. 5-17.

## Preserving the Pyramid of STI Using Buckets

Michael L. Nelson and Kurt Maly

### ABSTRACT

The product of research projects is information. Through the life cycle of a project, information comes from many sources and takes many forms. Traditionally, this body of information is summarized in a formal publication, typically a journal article. While formal publications enjoy the benefits of peer review and technical editing, they are also often compromises in media format and length. As such, we consider a formal publication to represent an abstract to a larger body of work: a pyramid of scientific and technical information (STI). While this abstract may be sufficient for some applications, an in-depth use or analysis is likely to require the supporting layers from the pyramid.

We have developed buckets to preserve this pyramid of STI. Buckets provide an archive- and protocol-independent container construct in which all related information objects can be logically grouped together, archived, and manipulated as a single object. Furthermore, buckets are active archival objects and can communicate with each other, people, or arbitrary network services. Buckets are an implementation of the Smart Object, Dumb Archive (SODA) DL model. In SODA, data objects are more important than the archives that hold them. Much of the functionality traditionally associated with archives is pushed down into the objects, such as enforcing terms and conditions, negotiating display, and content maintenance. In this paper, we discuss the motivation, design, and implication of bucket use in DLs with respect to grey literature.

1  
Digital  
Library

## INTRODUCTION

Research projects produce information in a variety of formats. A traditional formal publication, such as a journal article, is generally supported by a large quantity of software, datasets, images, video, informal notes, presentations and other documents. Collectively, we call this set of scientific and technical information (STI) the "Pyramid of STI".

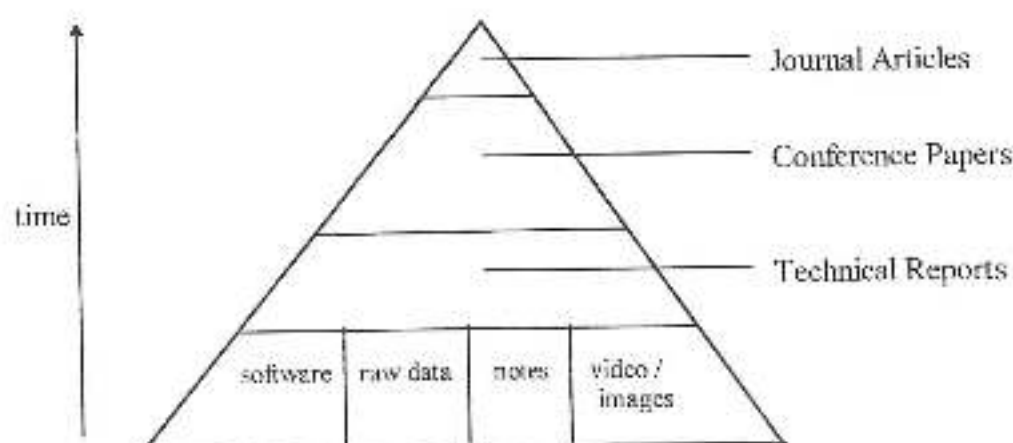


Figure 1: Formal Publications Rest on a Pyramid of STI

Although the information in this pyramid was created together and subtle relationships between its components can exist, different semantic instantiations are generally segregated along currently obsolete media boundaries. Reports are placed in report archives, software might go into a software archive, but most of the data, supporting materials and other grey information are likely to be kept in informal personal archives or discarded altogether. Our experience with NASA digital libraries (DLs) is users wish to have access to the supporting materials, data and software used in the preparation of formal literature (Sobieszczanski-Sobieski, 1994) -- even though there currently is no well established publication vector for much of this information.

We feel many DL projects focus on simply automating the formal publication process, while providing little attention to the lower tiers of the pyramid of STI. Similarly, we feel that creating "separate but equal" DLs implemented congruently with obsolete media boundaries pushes the burden of STI (re)integration to the user.



To address these concerns, we have created buckets: aggregative, intelligent agents tailored for publishing in digital libraries. Buckets can be thought of DL-specific Digital Objects as described in (Kahn and Wilensky, 1995). Buckets provide a container mechanism to capture and preserve the pyramid of STI. To enhance their long-term survivability, buckets are completely self-contained. [All the logic to manage, and protect their content is embedded in the bucket itself, not in a separate server.] Buckets can also be thought of as "archivelets". Communication with the bucket occurs through messages utilizing the bucket application programming interface (API). [By promoting buckets to first class network citizens, we are stressing that content is more important than the search engine or DL protocol used to access it.] By imbuing the content with additional functionality, we increase the long term usefulness and survivability of the STI.

### Grey Literature at NASA

NASA communicates its research findings through the traditional open literature process as well as its own multi-tiered, self-published report series (Pinelli, 1990). The NASA report series offers a number of advantages to authors: no page restrictions, potential for restricting dissemination, possibility of color graphics, and occasionally the inclusion of a CD-ROM of data, images or software. However, the latter two are rarer than most authors would like because they are expensive to create, and their distribution is more expensive still. The NASA reports are often ingested in systems that can handle only paper hard copy or possibly just microfiche -- leaving few options for propagation of additional media formats such as CD-ROMs.

An even more compelling case for capturing grey literature at NASA is that the formal publications (NASA's report series or open literature) represent a decreasing percentage of the total amount of STI created and used by NASA and its customers. [Due to the increasingly proprietary nature of NASA's work, as well as increasing time constraints on fewer staff members, many research projects are no longer resulting in a formal publication.] Instead, the projects remain as a collection of briefings, data, and other forms of grey literature -- often with proprietary access restrictions. While neglecting the formal publications achieves the short term goal of increased project turn around time, the inability to capture and preserve any

of the resultant STI creates a gap in the corporate memory. There is no well-defined, large scale publishing outlet for a majority of the STI created at NASA.

### Shortcomings of Current Digital Libraries

We recently surveyed a number of digital libraries from multiple disciplines and found that most focused only on formal publications (Eisler & Nelson, 1998). Even DLs that focus on grey literature (i.e. technical reports, pre-prints, working papers) still only focus on the hard-copy representation. Software, images, video and other material, if they are preserved at all, are often sent to DLs that serve only those respective media formats. Such examples include NIX, an image DL for NASA (von Ofenheim, et al., 1998); Nctlib, a software DL for the high performance computing community (Browne, et al., 1995); and Alexandria, a DL for geospatial data (Smith, 1996). While these and other like DLs can provide custom interfaces for interacting with non-textual STI, they have as a side effect an artificial segregation along media formats.

We are aware of no other current DL that allows for archiving and serving a complex set of STI. Consider a research project that produced raw data, software to reduce the data, reduced data, shift notes, image or video representation of data, presentations, informal reports, and a journal article that summarized the project's findings. Currently, the journal article has the best chance of appearing in a DL. Possibly the software will make it to a software DL, and depending on the discipline the data might make it to a data archive, but neither is guaranteed. The "base" of the pyramid of STI is at worst, effectively thrown away, and at best is splintered and sent to separate DLs. If researchers read the journal article and wish to extend the data analysis differently than presented, they can either search in other DLs for the software and datasets, or they can contact the authors in the hopes of obtaining the software and datasets through collegial distribution. The former assumes the various non-report STI has been placed in DLs somewhere, while the latter assumes the authors can still be reached.

### BUCKETS

- <sup>11</sup> Buckets are object-oriented container constructs in which logically grouped items can be collected, stored, and transported as a single unit. <sup>12</sup>Buckets are completely self-contained; they



have their content physically resident, the intelligence to serve, manage and enforce terms and conditions for accessing the content. Their self-contained nature allows them to function independently of, and in cooperation with, any DL system or protocol. It also allows them to be completely mobile - a bucket can "move around" and still retain its functionality. Buckets also maintain their own logs of actions performed on them, so they can retain a history that is independent of the servers used to access them.

As far as what goes into a bucket, buckets provide mechanism -- not policy. There are no pre-defined concepts of what information types should go into a bucket, and buckets make no assumptions about their content. The authors and publishing organizations control what constitutes a bucket; if there is reason why a lower strata in the pyramid should not be preserved, there is no requirement that it be included. Similarly, buckets are not unsuitable for white literature; they simply offer the capability (should the authors and publishers choose) to extend beyond the white into the grey.

- \* To the casual observer, a bucket appears as a regular web page. However, messages are sent to the bucket using the hypertext transfer protocol (http). If no message is sent, by default the bucket builds an HTML presentation page of its contents:

`http://dlib.cs.odu.edu/test-bucket3/`

However, other messages are possible. If a web robot wanted to gather the structured metadata, it could issue this bucket message:

`http://dlib.cs.odu.edu/test-bucket3/?method=metadata`

To learn what methods this bucket supports, it could issue:

`http://dlib.cs.odu.edu/test-bucket3/?method=list_methods`

Many other methods are defined, including methods listing the bucket source code, updating the bucket source code, changing who are privileged principals to the bucket, and other various functions. See (Nelson, et al., 1999) for a further discussion of bucket methods and implementation.



### SODA: SMART OBJECTS, DUMB ARCHIVES

An observation from our experiences with current NASA DLs is that a surprising number of people do not find the publications via the respective DLs. Since the full contents of the NASA DLs are browsable, both the abstract lists and the reports are indexed by web crawlers, spiders and the like. Users are formulating complex queries to services such as Yahoo, Altavista, Lycos, Infoseek, etc. to find NASA STL. We presume this is indicative of the resource discovery problem: people start at these portals because they do not know all the various DLs themselves; and the meta-searching problem: they are trusting these services to search many sources, not just the holdings of a single DL.

Although we believe we have built attractive and useful interfaces for the NASA DLs, our main concern is that people have access to NASA's holdings and not that they use a given DL interface. It is desirable that NASA publications are indexed by many services. Since there can be any number of paths to the information object, the information object must be a first class network citizen, handling presentation, terms and conditions, and not depending on archive functionality.

In the SODA model, we have separated DLs into three separate layers: 1) the Digital Library Services (DLS) layer that provides user interfaces such as searching and browsing; 2) the archive layer that manages collections of objects; and 3) the objects themselves. Separating the functionality of the archive from that of the DLS allows for greater interoperability and federation of DLs. The archive's purpose is to provide DLs the location of buckets (the DLs can poll the buckets themselves for their metadata), and the DLs build their own indexes. And if a bucket does not "want" to share its metadata (or contents) with certain DLs or users, its terms and conditions will prevent this from occurring. For example, we expect the NASA digital publishing model to begin with technical publications, after passing through their respective internal quality control, to be placed in a NASA archive. The NASA DL (which is the set of the NASA buckets, the NASA archive(s), the NASA DLS, and the user communities at each level) would poll this archive to learn the location of buckets published within the last week. The NASA DL could then contact those buckets, requesting their metadata. Other DLs could index NASA holdings in a similar way: polling the NASA archive and contacting the appropriate buckets. The buckets would still be stored at NASA.

but they could be indexed by any number of DLs, each with the possibility for novel and unique methods for searching or browsing. Or perhaps the DL collects all the metadata, then performs additional filtering to determine applicability for inclusion into their DL. In addition to an archive's holdings being represented in many DLs, a DL could contain the holdings of many archives. If we view all digitally available publications as a universal corpus, then this corpus could be represented in  $N$  archives and  $M$  DLs, with each DL customized in function and holdings to the needs of its user base. The SODA model for DLs is discussed in detail in (Maly, et al., 1999a).

### FUTURE WORK

There are several areas we continue to explore and develop, the first of which is tools for buckets. The long-term success of buckets will depend on the quality of the tools to create and manage buckets. We have Publishing Tool, that allows users to transparently create and populate buckets. We have a Management Tool, which is a simple workflow mechanism to review buckets submitted for publication, approve or reject them, and move them into designated areas when approved. Finally, we have an Administration Tool that allows for long term maintenance of buckets, performing such operations as large scale source code and principal updates. These tools are undergoing frequent revisions as we gain user feedback.

Secondly, we are developing the concept of discipline specific buckets that can exploit knowledge about their contents. These bucket templates could have specialized display capabilities, specialized methods for data interaction, custom storage options, etc. These would be in contrast to the ontology neutral buckets presented here. The first of these, a bucket template for undergraduate education, is discussed in (Maly, et al., 1999b).

The third area we are actively developing is making buckets intelligent agents. The buckets are already carrying the intelligence to manage their content -- why not make them even more intelligent? We are designing a Bucket Communication Space (BCS) that will allow buckets to communicate with other buckets, people, and arbitrary third party network resources. The BCS will allow for functions such as messaging, format conversion, and matching. The latter is especially interesting: we will allow buckets to find similar buckets off-line. Buckets will register their profiles with the BCS, and the BCS will inform buckets of their potential matches



based on a similarity index. Buckets can then either automatically establish linkages, or contact their owners for verification.

Additional functions afforded by intelligent buckets includes a redefinition of software reuse. If general software resides in a bucket in a DL, users can either download and integrate it with their applications, or use the software while it is resident in DL. Many DL projects have agent technology designed to assist the user in searching, browsing or other functionality. However, we are unaware of other projects that attempts to make the archived object itself intelligent. The full implications of making archival objects intelligent agents has yet to be explored.

## CONCLUSIONS

Past media format limitations have defined our view of an archival unit of STI. We claim the formal publications that are the focus of traditional libraries and many DL projects are simply an abstract to a larger body of information. This pyramid of STI is supported by informal documents, software, data, images, videos and other multi-format material. Our DL experiences have lead us to believe that there is high user interest in obtaining access to the lower tiers of the pyramid of STI — many of which are not archived at all.

To this end, we have developed buckets: object-oriented, intelligent agents for publishing in DLs. Buckets are DL system and protocol neutral; they depend on no particular system and are designed to have minimal impact on any such system. Messages are communicated to buckets via an API that uses http as a transport. To increase their long term survivability, buckets are completely self-contained, carrying all of their contents internally as well as the logic to manage, serve and protect those contents. In addition to the aggregative properties of buckets, we are also working to make them intelligent agents. We are implementing the Bucket Communication Space to allow buckets to communicate with each other and perform tasks such as finding and linking to similar buckets. By imbuing STI objects with aggregation and intelligence, we can now preserve the entire pyramid of STI, as well as introduce new functional capabilities for archived objects.



## REFERENCES

- Browne, S.; Dongarra, J.; Grosse, E. and Rowan, T. (1995), "The Netlib Mathematical Software Repository", D-Lib Magazine, September 1995. Available:  
<http://www.dlib.org/dlib/september95/netlib/09browne.html>
- Esler, S. L. and Nelson, M. L. (1998), "Evolution of Scientific and Technical Information Distribution", *Journal of the American Society of Information Science*, Vol 49 No 1, pp. 82-91.
- Kahn, R. and Wilensky, R. (1995), "A Framework for Distributed Digital Object Services", *cnri.dlib/tn95-01*. Available:  
<http://www.cnri.reston.va.us/ctr/arch/k-w.html>
- Maly, K.; Nelson, M. L. and Zubair, M. (1999), "Smart Objects and Dumb Archives: A User-Centric, Layered Digital Library Framework", D-Lib Magazine, Vol 5 No 3. Available:  
<http://www.dlib.org/dlib/march99/maly/03maly.html>
- Maly, K.; Zubair, M.; Liu, X.; Nelson, M. L.; Zeil, S. J. (1999), "Structured Course Objects in a Digital Library", to appear in *Proceedings of the International Symposium on Digital Libraries*, Tsuka Japan.
- Nelson, M. L.; Maly, K.; Zubair, M. and Shen, S. N. T. (1999), "Buckets: Aggregative, Intelligent Agents for Publishing", *Webrnet Journal*, Vol 1 No 1, pp. 58-66.
- Pinelli, T. E. (1990), "Introduction to National Aeronautics and Space Administration's Scientific and Technical Information Program", *Government Information Quarterly*, Vol 7 No 2, pp. 123-126.
- Smith, T. (1996), "A Brief Update on the Alexandria Digital Library Project", *D-Lib Magazine*, March 1996. Available:  
<http://www.dlib.org/dlib/march96/briefings/smith/03smith.html>

Sobieszczanski-Sobieski, J. (1994), "How to Improve NASA-Developed Computer Programs", NASA CP-10159, pp. 58-61.

von Ofenheim, W. H. C.; Helmreich, N. L.; Binkley, R. L.; Curry, M. A.; Slater, R. T.; Nolan, G. J.; Griswold, T. B.; Kovach, R. D.; Corbin, B. H. and Hewitt, R. W. (1998), "NASA Image eXchange (NIX)", NASA/TM-1998-206295.

Michael L. Nelson  
 NASA Langley Research Center  
 MS 158  
 Hampton, VA USA 23681  
 +1 757 864 8511  
 +1 757 864 8342 (fax)  
 m.l.nelson@larc.nasa.gov  
<http://home.larc.nasa.gov/~mln/>

Michael is an electronics engineer at NASA Langley Research Center and a Ph.D. candidate in computer science at Old Dominion University.

Kurt Maly  
 Department of Computer Science  
 Old Dominion University  
 Norfolk, VA USA 23592  
 +1 757 683 4817  
 +1 757 683 4900 (fax)  
 maly@cs.odu.edu  
<http://www.cs.odu.edu/~maly/>

Kurt is the Kaufman Professor and Chair of the Computer Science Department of Old Dominion University.

## AN ARCHITECTURE FOR GREY LITERATURE IN A R&amp;D CONTEXT

Prof. Keith G Jeffery

Acting Director Information Technology

CLRC-Rutherford Appleton Laboratory

Chilton, Didcot, OXON OX11 0QX UK

(keith.g.jeffery@rl.ac.uk)

**Abstract**

The importance of Grey Literature is becoming increasingly recognised. For many organisations it encapsulates the knowledge and know-how and thus is a vital business asset. It has similar importance in quality of life aspects - healthcare, environment and culture. Grey Literature in a R&D environment represents the cutting edge of this knowledge and so its management is of utmost importance. Partly based on involvement in defining the datamodels for R&D information interchange across Europe, the author here defines a content metadata datamodel for Grey literature which is more expressive and has more flexibility than any previous proposal and which integrates seamlessly with the CERIF2000 definition which will soon replace the CERIF1991 European Union Recommendation to Member States on Exchange of R&D Information. The content metadata datamodel offers significant advantages over Dublin Core yet can generate Dublin Core if required.

**Keywords:**

grey literature, metadata, W3C, catalog, digital library, CERIF, ERGO, datamodel, workflow, Dublin Core



## INTRODUCTION

### R&D Information

It is generally acknowledged that easy access to R&D information provides innovation motivation for industry and commerce leading to wealth creation and employment, and simultaneously provides quality of life advantages especially in healthcare, environmental and cultural aspects. Provision of such access is particularly interesting in a multilingual, multicultural environment such as Europe.

Work on many aspects of R&D information has been going on for many years, and systems have been developed and are in use. Integration of such systems has been an interest of the author for many years. Now that - through the CERIF and ERGO initiatives - integration of information on R&D projects across Europe is well underway, and integration of information for persons and organisations planned together with links to patents, products and publications, it is time to consider how best to address Grey Literature. Some attempts have been made to integrate Grey Literature into conventional library electronic catalogue systems e.g. (BibSys) and a particular catalog format was adopted for SIGLE (Farace, 1990). The system for handling theses at Virginia Tech (Virginia) has defined its own metadata including that for controlling workflow. However, these attempts do not satisfy the requirements identified.

It is becoming increasingly clear that in many organisations the Grey Literature documents as information, the knowledge and know-how of the organisation - assets that are extremely valuable - and therefore which need to be curated carefully and used for trade with other organisations subject to security and access controls. In a R&D environment such knowledge is at the cutting edge of progress and is thus extremely valuable.

The end-user requirement is for the relevant information (relevance, recall), at the right place (wherever worldwide), at the right time (when required), in the appropriate form (optimal presentation, integrated for further use in electronic information / office environments). The end-user requires this information to be available through an easy-to use, tolerant, intelligently-assisting, multilingual, multimedia, multimodal user interface. Briefly, an interface that obeys 'do what I need, not what I say'.

### History of the Paper

The author has designed R&D systems for UK Research Councils and currently leads the project integrating UK R&D resources for presentation to common EU standards. He was involved in the definition of CERIF1991. He is currently designing a system for Grey Literature at CLRC-RAL, having experimented with various digital library environments including within the (DELOS) project over many years. The author has also been associated with work on metadata for R&D projects within the (FORSKDOK) project at University of Bergen for some time. Grey literature has been recorded in the (BibSys) System with limited success.

The author represents UK at the European Union CERIF (Common European Research Information Format) Revision Working Group where the topic of Grey Literature has yet to be addressed - although most other aspects of R&D information have already been discussed and standards agreed - largely based on the datamodels proposed by the author. Thus in the second half of 1998 the author decided to work intensively on Grey Literature; because of the ongoing cooperation with the team there led by Jostein Hauge, a Workshop was organised in Bergen in December 1998. An early version of some of the ideas in this paper was presented there.

### Structure

The rest of the paper is organised as follows: Section 2 reviews the background state of R&D information concluding with the urgent need for systems for Grey Literature. Section 3 discusses requirements of a Grey Literature System and Section 4 develops an architecture. Section 5 discusses options in implementation and presents a datamodel for Grey Literature Content metadata which is a superset of other content metadata standards and proposals. Current work on other kinds of metadata is outlined. Section 6 concludes.

### BACKGROUND : R&D SYSTEMS

The Grey Literature requirement here is viewed as one component of a much larger environment for R&D (Research and Development) information. This implies that a Grey Literature Resource (object) will have relationships with persons, organisational units, projects, patents, products, other publications.

### The Customers / Users

The customers / users of a R&D system and their particular requirements include:

- (a) Researchers: find partners, locate techniques, manage projects, produce outputs;
- (b) University Officials: authorise, examine, produce statistics;
- (c) Research Council (R&D Government funding source) officials : evaluate projects and results, produce statistics;
- (d) Commerce & Industry: find ideas, technologies of potential relevance;
- (e) Intermediaries: assist commerce and industry in finding relevant innovative technologies;
- (f) Media: find news articles;
- (g) General Public: general awareness of R&D.

### The Products of R&D

The products of R&D are listed below. Items in **bold** are regarded as 'grey literature': there is a need to design to manage all of these kinds but it was decided to concentrate on publications first:

- (a) publications: journal / conference papers, **technical papers, theses, dissertations, reports**
- (b) patents
- (c) products: **prototypes**, fully engineered products
- (d) results: **data and its presentation / visualisation**
- (e) know-how and IPR: **reports, procedure instructions**
- (f) education and training: **documentation, courses**
- (g) publicity: **press releases, product or organisational 'flyers'**



### R&D Project

The products are the output from a project, which has a mission, objectives, deliverables, project plan, a duration and resources. The project has a need for access to pre-existing and current information. The project goes through the stages: proposal / project execution / evaluation. Additionally, Researchers need project management / accounting tools but these are not considered here.

### Pre-Existing R&D Systems

It is pointless, both for legacy-usage and economic reasons, to re-develop in a R&D environment systems for patents, products, publications; they exist and they work. Of course there are shortcomings and criticisms but in general the systems provide a service e.g library catalog systems. Hopefully they will evolve to continue to meet user requirements, if not replacement will have to be judged on the business case at the time.

However, it should be noted that existing systems are distributed, sometimes unconnected and heterogeneous in the following characteristics:

- (a) purpose and intent;
- (b) structure (relationships);
- (c) content (attributes, language, character set);
- (d) available input / output interfaces;

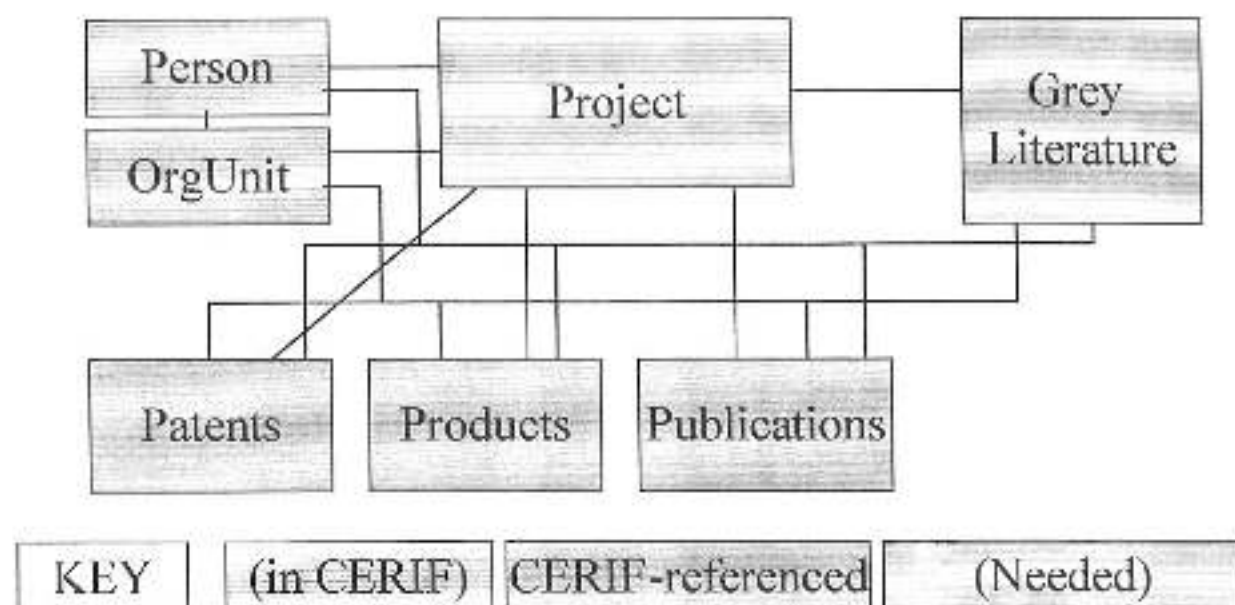
so it is necessary to provide interoperability interfaces in order for the end-user to have an integrated information landscape. This is achieved using metadata and associated processing.

Pre-existing systems for R&D projects also exist; they, too, are distributed and heterogeneous and the ERGO project utilising the CERIF1991 Recommendation is integrating across this heterogeneity using an enhancement of the the catalog technique (Jeffery, 1992) as used in IDEAS (Jeffery et al, 1989) and the subsequent EXIRPTS projects - thus providing a common datamodel. The CERIF2000 recommendations formalise and improve the datamodels considerably.

A similar technique is proposed here for Grey Literature, to provide through a common catalog (based on a formal metadata datamodel), easy-to-use access to heterogeneous sources of grey literature.

### CERIF

CERIF provides a comprehensive data model for R&D Information agreed by representatives of European countries (both European Union and associated States). The original CERIF1991 recommendation is being used as the basis for the ERGO Pilot Project (ERGO). The new CERIF data model has been produced within the CERIF Revision Working Group, dominantly by the author. The new CERIF datamodel is being implemented in UK as a RDBMS to ensure completeness, integrity and requirements satisfaction. CERIF assumes pre-existence of systems for patents, products and publications. However, CERIF has not (yet) tackled the problem of grey literature - hence the work by the author presented here. CERIF may be represented (Figure 1: CERIF High-Level Datamodel):



**Figure 1: CERIF High-Level Datamodel**

and from this 'ideal CRIS' (Current Research Information System) datamodel- which could be implemented in whole or part by anyone wishing to set up a R&D database system - a subset for exchange (respecting e.g. privacy legislation or commercial confidence) is derived and also a

datamodel for content metadata for use in a catalog. The full ERGO project architecture envisages end-user query to the catalog, from the 'hits' choice of relevant resources (objects) from the host databases and their provision via the exchange datamodel to the end-user as hypermedia documents.

## **GREY LITERATURE SYSTEM**

### **Grev Literature System Requirements**

The major requirements are as follows:

- (a) solicit / accept 'documents' produced in a variety of electronic forms (hyperlinked multimedia);
- (b) convert them to a canonical form for storage, including - where required - enforcing style (or store in original form with metadata to cause conversion);
- (c) register them and catalogue them in a suitable way;
- (d) provide suitable IPR, copyright and other legal protection;
- (e) make provision for charging for end-user services by end-user category and usage statistics collection;
- (f) provide a system for information retrieval over the documents (including multimedia);
- (g) make them available (subject to security / authorisation) to end-users in a variety of forms;
- (h) provide workflowed input / update

The 'documents' are:

- (a) as input: atomic documents / document fragments or hyperlinked document sets;
- (b) as output: atomic documents / document fragments or hyperlinked document sets

The key question then is what is the atomic unit? On input it is likely to be a complete document (author of complete opus) whereas on output it is likely to be the relevant fragment



of the document hyperlinked with other fragments of this and other documents to satisfy the user requirement.

Therefore it is essential to separate clearly:

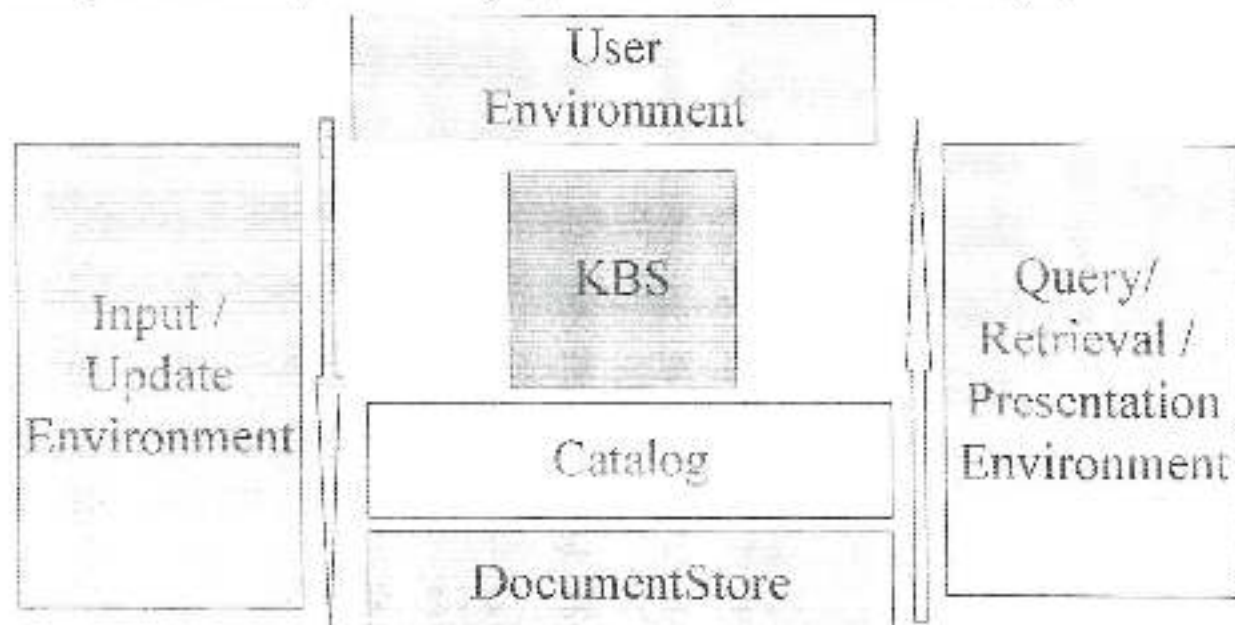
- (a) document structure (hyperlinks between document atomic fragments);
- (b) document content (attributes or sections, language, character set, media representation(type));
- (c) document presentation / layout (for preservation of presentational features);
- (d) document metadata (e.g. cataloguing information, security / access, charging, provenance, certification..)

A Grey Literature Document is likely to be of text, graphics, and images and may include video and audio - but also with experimental results, virtual reality presentations - thus provision of 'drill-down' from the conventional view of a document to its source data will be essential.

## **GREY LITERATURE SYSTEM ARCHITECTURE**

### **Simple Architecture**

A simple architecture (where 'catalog' represents all required kinds of metadata) is:



**Figure 2: Simple Architecture**

### Major Architectural Ideas

The major architectural ideas are as follows:

- (a) accept and provide multiple forms - different formats (SGML, XML, HTML, Word...) and full multimedia (SMIL);
- (b) the use of rich metadata (of different kinds) not only as an index to the document content but also as a mechanism for security, pricing control, access information (royalty payments etc), provenance, certification - all linked to the Grey Literature processes implemented as workflow;
- (c) using KB assists and style controls to propose metadata catalogue entries having parsed the document or check input for quality;
- (d) the KB-assist knowing the metadata catalog and the user profile (implies domain ontology with the KBS) refines queries, improves integration of the answer - with appropriate explanation and controls presentation style for end-user preferences.

### Architectural Features - Input / Update Forms and User Interface

#### The Input /Update Forms

The major idea is to use standard WWW HTML forms. However, using RDF-Workflow Technology (from the W3C-LA (World Wide Web Consortium - Leveraging Action) project (W3C-LA)) we can ensure that suitable authorisations occur for document placement in the library - e.g. author signature, supervisor signature, department signature and all with DSIG (Digital Signature) technology (for security) if required.

#### User Interface

The User Interface has to integrate at user client workstation access to all the required information sources, for both output and input as authorised, at the desk or in a mobile environment. It has to operate on various hardware / operating system platforms (Unix workstation, PC, mobile phone with display). There is only one solution for the architecture: use of a WWW Browser.

The browser should be lightweight and with a very fast core, and allowing add-ons and plug-ins configured dynamically as required. Experiments with such a browser have been performed at CLRC-RAI (for another project).

The browser should be supported by an easy-to use, tolerant, intelligently-assisting, multilingual, multimedia, multimodal user interface, characterised by:

- (a) easy to use : intuitive, minimal interaction;
- (b) tolerant : accepts incomplete and uncertain data (input and output) and proposes supplements to completion;
- (c) intelligently-assisting : query expansion / refinement - dialogue with user and answer expansion / refinement / explanation;
- (d) multi-lingual, multimedia, multimodal: all forms of human-computer interaction supported

Such an interface is to be implemented using Java Applets so that central control may be maintained, linked with evolving domain ontologies and multilingual thesauri, and compatibility with all reasonably modern browsers maintained.

### **Architectural Features - Metadata**

There is relevant information on digital library metadata at the (UKOLN) site including pointers to the current standards and state of the art.

#### **Content Metadata : The Catalog**

This kind of metadata is equivalent to the conventional library card catalogue. However,

- (a) we have defined a content metadata database structure that is much more flexible than e.g. Dublin Core (DC) and is defined formally (Jeffery, Asserson, 1998a);
- (b) the design will use a structured DBMS (Database Management System) rather than an IR (Information Retrieval) system for fast retrieval, standardisation of query languages and optimal concurrency control, but especially to allow for flexible structural relationships both within content metadata and between content metadata and other kinds of metadata;



(c) the design provides for interoperability by providing interconversion from the catalog metadata to conventional Dublin Core stored in RDF (Resource Description Framework, the W3C metadata standard (RDF)); to multiple Dublin core extensions stored in RDF and to MARC (MARC) if required (as already demonstrated by the Nordic Metadata Project (NMP)) or to other metadata formats. This provides for maximum interoperability with other digital library projects yet does not constrain this design like the others.

### Structure Metadata

The need to separate structure from content is well-known in the Hypermedia community by the use of a linkbase e.g. Hyperwave (HW), Microcosm (MI). Unfortunately HTML was not designed with this in mind but mixes within the document hyperlinks (structure) and content.

However, there is a clear requirement to extend beyond the simple linkbase concept to accommodate links with semantics, ideally expressed as first order logic semantics with second order syntax (complex structures). This requirement and a solution has been demonstrated in interoperation of hyperlinked multimedia clinical patient records in the project Hypermetadata (HMD). Such rich structure metadata implies deductive / inductive capability in the system - and thus the requirement for a KBS. For Grey Literature, as a first step, the use of extended RDBMS (Relational DBMS) technology is sufficient and provides an upgrade path to a richer system.

### Presentation Metadata

This provides the recommended presentation of a complete document or document fragment for the appropriate stored form. Examples are : SGML - DSSSL, XML - XSL, WORD - Stylesheet, full hypermedia - SMIL. Alternatively, documents can be stored in page-resolved form to preserve exactly the presentation required. Examples are PDF or Postscript.

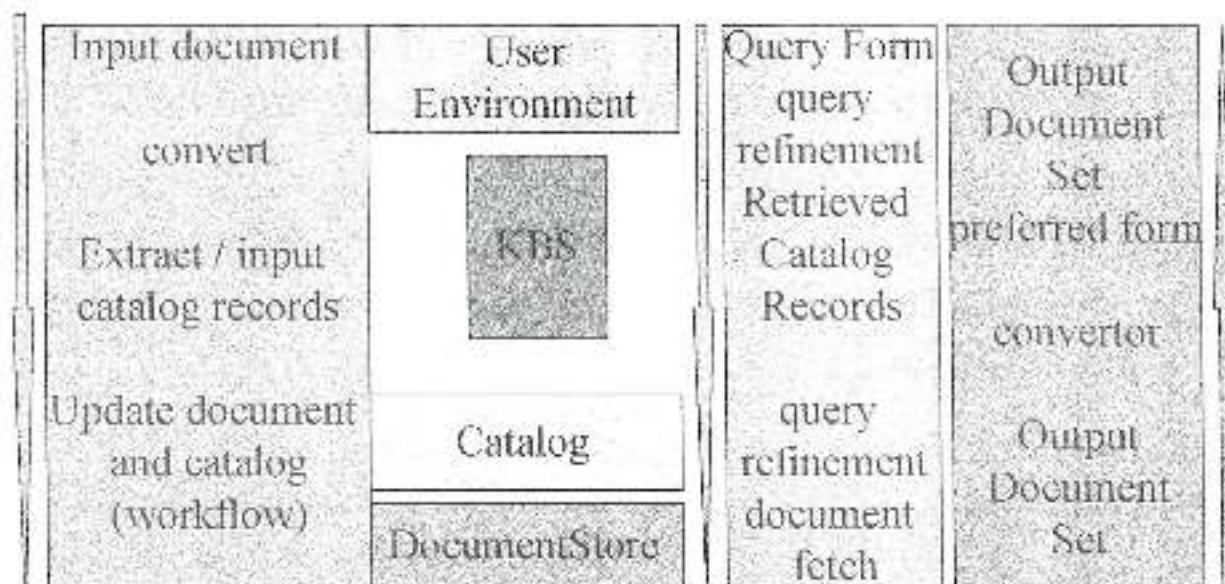
### Security / Access + Charging / Workflow Metadata

This kind of metadata provides appropriate information for processes associated with Security, privacy, authorisation, (copyright), with access records (charging / royalties), with document provenance and history - versions (link to workflow) and with document certification (e.g.

signed by author as original, signed by head of department as approved, signed by publication board as refereed (all linked to workflow).

### Architecture - Detail

The above features lead to a more complex architecture (Figure 3: Detailed Architecture) while still preserving the essential features of the simple architecture (Figure 2: Simple Architecture).



**Figure 3: Detailed Architecture**

### Grey Literature System - Implementation

The architecture has been designed specifically so that it is possible to start simply and grow to the KBS assisted architecture starting with assists for the user interface and queries then to provide assistance in results and presentation. The implementation options have been left as flexible as possible: for example the KBS could be implemented as a central system (as in project MIPS (MIPS)) or as cooperating agents e.g. as used at (Stanford).

The local process environment will dictate the need for workflow for authorisation, and for converters to handle different input and output forms in the environments for input / update (with workflow) and retrieval / presentation.



## **DISCUSSION**

Implementing such an architecture raises several interesting implementation decisions which are discussed here.

### **Storage & Retrieval**

- (a) centralised or distributed : multimedia assets are likely to be most accessed locally and are expensive to transmit so distributed is the best model. In practice logical central design and distributed physical design is ideal, but due to pre-existing systems unrealisable;
- (b) filestore or database store : the criteria are performance, security, and integrity. Database technology will be better for the latter two, performance will depend on several factors although there is some evidence that filestore systems perform slightly better than DBMS for Grey Literature Resources;
- (c) IR or Structured DB : the criteria are retrieval capability, update, and flexibility of structural linkage. A structured database - especially extended relational - will be better for almost all aspects, however, the one area where IR systems still excel is the use of inverted indexes for full-text searching. On balance, the advantages of Extended RDBMS outweigh those of IR systems.
- (d) same storage for metadata and data (objects, documents): there are advantages if both the metadata and data share a common environment for performance, security, integrity and a common DDL (data definition language - language to define schema) and common DML (data manipulation language eg SQL).

### **Presentation**

The choice of presentation environments - and metadata structures to support them - is critically important not only for aesthetics but also for ensuring certain resources are preserved as presentationally intended for legal reasons.

The SGML Family: SGML: DTD, SGML, DSSL, XML: DTD, XML, XSL and, of course SMIL provide matching content/structure and presentational capabilities and are clearly the best



choice with the richest (SGML) as canonical storage form; in practice it is likely that XML will be more widely used and this may well prove the best choice.

HTML: (DTD), HTML4.0, CSS2 provide some separation of content/structure from presentation but not separation of content and structure. Where it is not possible to solicit input in the SGML family or present using the SGML family then the HTML family is a widely available lower standard alternative.

However, there are many users of the Microsoft family of products: WORD: StyleSheet. However, the Stylesheet has rather limited presentational expression and so conversion from WORD to a canonical form is preferable.

Finally, there is the PageFamily: PDF, Postscript, proprietary DTP formats. Postscript should be used only as an interface between server and printer - there are too many variants for real compatibility. PDF provides a reasonably compact 'page image' form and is useful for preservation of intended presentation.

### Input

There are advantages for cross-platform compatibility and maintenance (including downline loading of Java Applets) in using for input / update Web Forms. The optimal design decisions appear to be to handle the data so collected as follows: metadata --> XML; content--> XML or SGML or SMIL;

Alternatively, especially for large volumes of data it may be better to use a Context Sensitive Editor. In this case similar design decisions apply: metadata --> XML; content --> XML, SGML or SMIL.

However, if a proprietary Word processor system is used e.g. Word, or a Text Processor e.g. LATEX or a proprietary DTP System then it will be necessary to apply converters - with success depending on the capabilities of the conversion software and the capabilities of the data generation systems to provide rich content and structure.

### Metadata

One major architectural decision is that the key to success is metadata: it is used for

- (a) quality input : domain range checking, attribute value confirmation;
- (b) workflowed input process (including review, authorisation): ensuring integrity;
- (c) security, access control, charging, royalties: preserving the social and economic aspects of the system;
- (d) query assistance / optimisation : by use of the KBS interpreting the metadata of the domain ontology to improve the query formulation and to assist optimisation;
- (e) answer assistance / optimisation : by use of the KBS using metadata concerning presentation;
- (f) interoperability across heterogeneous sources : by use of the KBS and domain ontology providing schema reconciliation with metadata assistance in term clarification and domain comparison;
- (g) user interface adjustment / assistance : using metadata concerning the user profile and preferences;
- (h) help and explanation - education : using metadata to explain data content and structure, presentation and quality, and to explain process (such as workflowed) with associated rules and conditions

An overview of Metadata for a R&D environment was presented at the CRIS98 Conference (Jeffery, 1998). W3C working groups are finalising RDF with XML implementation at the time of writing. Separately, the Dublin Core work appears to have stalled due to differences of opinion within the working group(s).

We have reviewed carefully content and structure metadata (that required for description of content and approximately equivalent to a library catalogue card) and have proposed a metadata datamodel which is a superset of all other content models and so they can be generated from it.

The datamodel is formal and yet flexible. Furthermore, it is extremely suitable for Grey Literature in a R&D environment because it integrates seamlessly with CERIF.

A current solution should be implemented on extended RDBMS technology since it provides adequate, if not extensive, intensional (schema) capability and sufficient formality. It also provides great flexibility for evolution towards more expressive environments (e.g deductive / inductive). The content metadata is detailed; other kinds of metadata are sketched and await further work in parallel with the work of W3C.

Our work leads to the following datamodel: (compared with the Dublin Core, deleted elements in *italics*, new elements in **bold**, all elements redefined)

content metadata:

**UniqueId**, Title, *Author or Creator*, **Person**, **OrgUnit**, *Subject and Keywords*, **Subject**, **Keywords**, Description, *Publisher*, *Other Contributor*, Date, Resource Type, Format, Resource Identifier, *Source*, *Language*, *Relation*, *Coverage*, **CoverageTemporal**, **CoverageSpatial**, *Rights Management*

note: each textual field has subelements language, representation, format; many elements have schemes; any element may be repeated. Thus this is normalised to:

<UniqueId> <sup>/\*</sup><Resource Identifier> where the value of Resource Identifier is a pointer (typically a URL). /\* links the URL or other pointer indicating the real resource (eg document) to a Unique Id in the metadata \*/

<Unique Id> <relation> <Person> where sub-elements of relation are type (meaning role), degree (cardinality) with mandatory / optional usage; associated relation constraints. Such relationships (roles) should have amongst their constraints temporal constraints including date or period. Sub-elements of person are a full attribute set from CERIF. Such relationships (roles) should have amongst their constraints temporal constraints including date or period.

<Unique Id> <relation> <OrgUnit> where sub-elements of relation are type (meaning role), degree (cardinality) with mandatory / optional usage; associated relation constraints. Such relationships (roles) should have amongst their constraints temporal constraints including date or period. Sub-elements of OrgUnit are a full attribute set from CERIF.



privacy metadata:

<Unique Id> <Privacy Scheme> <Privacy level constraints>

access metadata:

<Unique Id> <Access Rights Scheme> <Access level constraints>

charging metadata:

<Unique Id> <Charging Scheme> <Charge>

annotation metadata:

<Unique Id> <relation> <Annotation> <Person>

where <Annotation> has sub-elements language, representation, format. It is commonly necessary for comments upon a resource (annotations) to be recorded. The idea would be to define the purpose of the annotation by the type of relation (sub-elements of relation are type (meaning role), degree (cardinality) with mandatory / optional usage; associated relation constraints). Such relationships (roles) should have amongst their constraints temporal constraints including date or period) and for the annotator to be identified. In this case a DSig (digital signature) sub-element may be necessary under Person. This needs more discussion to determine the functional requirement.

These latter (non-content) kinds of metadata require further work (currently ongoing) to establish correctly the functional requirement, and then compare with emerging results of work elsewhere e.g. W3C. The approach taken here is much closer to the RDF (encoded as XML) approach of W3C - where essentially everything is reduced to binary relationships in a simple graph object model - than the existing Dublin Core work. A relational model of content metadata has been prepared and documented (Jeffery, Asserson, 1998a).

## CONCLUSION

A joint project involving University of Bergen, University of Oslo (Norway) and CLRC-RAL (UK) has been initiated, following the Bergen Grey Literature Workshop, December 1998 (Be98) with the major primary aim of providing a Grey Literature System for research theses.

## **ACKNOWLEDGEMENTS**

The author wishes to acknowledge the work of colleagues in the CERIF Revision Working Group which provided part of the motivation for this work. The author also wishes to acknowledge the work of colleagues at the Bergen Grey Literature Workshop, December 1998, especially the organiser, Anne Asserson (who prompted this piece of work) and the team from University of Oslo led by Hana Konupck.

## REFERENCES

(Be98) [http://www.fou.uib.no/fou/grey\\_lit/](http://www.fou.uib.no/fou/grey_lit/)

(BibSys) <http://www.bibsys.no/english.html>

(DELOS) <http://www.iei.pi.cnr.it/DELOS/>

(DC) [http://www.oclc.org/5046/research/dublin\\_core/](http://www.oclc.org/5046/research/dublin_core/)

(ERGO) <http://www.cordis.lu/ergo/>

(Farace,1990) SIGLE - System for information on grey literature in Europe Farace, D.L. In: Login, vol. 14, no. 3, p. 6-9, 1990 - ISSN0920-0290

(FORSKDOK) <http://www.fou.uib.no/eng/>

(HMD) <http://www.itd.clrc.ac.uk/Activity/Hypermedata>

(HW) <http://www.hyperwave.de/>

(Jeffery,1992) Jeffery,K.G.'Database Integration' Proceedings SOFSEM92 Conference, Ed. M Bartosek, Masaryk University Brno, Czech Republic

(Jeffery, 1998) <http://www.cordis.lu/cybercafe/src/jeffery.htm>

(Jeffery, Asserson,1998a) [http://www.fou.uib.no/fou/grey\\_lit/literature.html](http://www.fou.uib.no/fou/grey_lit/literature.html) (click on Data Model)

(Jeffery et al,1989) Jeffery,K.G.;Lay,J.O.;Miquel,J.F.;Zardan,S.;Naldi, F.; Vannini-Parenti,I.: 'IDEAS: A System for International Data Exchange and Access for Science'. Information Processing and Management 25(6) pp 703-711. 1989.

(MARC) <http://leweb.loc.gov/marc/marc.html>

(MI) <http://www.mmrg.ecs.soton.ac.uk/projects/microcosm.html>

(MIPS) <http://www.itd.clrc.ac.uk/Activity/MIPS>

(NMP) <http://renki.lib.helsinki.fi/meta/>

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

(Stanford) <http://walrus.stanford.edu/diglib/>

(UKOLN) <http://ukoln.bath.ac.uk/metadata/>

(Virginia) <http://www.ndltd.org/>

(W3C) <http://www.w3.org>

(W3C-LA) <http://www.itd.clrc.ac.uk/Activity/W3C-LA>



## Grey Literature in Energy: A Shifting Paradigm

by Deborah E. Cutler, U.S. DOE/OSTI

There were challenges to undertake and exciting new partnerships to forge as the exchange medium for energy information moved from paper and microfiche to electronic formats. As the paradigm for storing, retrieving, and disseminating energy-related grey literature has shifted and evolved, the Office of Scientific and Technical Information (OSTI), the Department of Energy's (DOE) information arm, has led the way. At this particular stage in the shift, OSTI has overcome a number of challenges and recognizes that there are still more to come. Dealing successfully with those challenges has come only as a result of a commitment to vision and cooperation/support from domestic and international partners. Of course the vehicles provided by the Internet, the World Wide Web, and related technologies were key, as well. OSTI has been able to offer the user community direct desktop access to a significant volume of energy-related grey literature. That access is FREE.

Producers and users of information have found themselves in a rapidly changing environment as the Internet and its support technologies have exploded. A companion event to that technology explosion has been the slashing of government budgets and dramatic downsizing of the government workforce. Government agencies can no longer afford to wait 10 years for proven solutions from private industry. There is an immediate need for answers to critical questions. How do you do more with less? What is the most efficient way to meet and exceed customer expectation? How can you keep commitments? How do you remove the "grey" from the traditional definition of grey literature? OSTI learned the answer to "how"—you take risks, you jump on the bandwagon, you just do it.

OSTI has been in the forefront and become a leader among US government agencies in offering grey literature via the Web. Evolving from the traditional world of paper and microfiche to the electronic world of PDFs and TIFFs has come quickly and not without growing pains. This paper documents OSTI's journey toward that new frontier and the challenges faced along the way.

### THE OLD AND NEW FRONTIER

Although the name and chain of command have shifted several times over the years, OSTI's core mission to collect, preserve, and disseminate scientific and technical information (STI) has remained the same. Over the last 52 years, OSTI's collection of STI has grown to over 1.5 million Energy R&D reports (traditionally viewed as grey literature) and approximately 5 million bibliographic citations representing worldwide energy R&D. The current database, today known as the *Energy Science and Technology Database (EEDB)*, has over 3.8 million citations. Over the past few years, OSTI has made significant strides into the information age: defining new electronic exchange formats; creating and linking to collections of digitized STI;

serving researchers directly; and developing a Web product called EnergyFiles: Virtual Library of Energy Science and Technology. As of September 1999, the DOE Information Bridge, a component of EnergyFiles, has over 56,000 full-text R&D grey literature reports (approximately 3.8 million pages), capturing a significant portion of DOE and global R&D output since 1996. The global information comes in through OSTI's partners in international agreements. To keep to the terms of these agreements, access to the full DOE Information Bridge (<https://apollo.osti.gov/dds>) is limited to the DOE and its contractor community. There is, however, public access to DOE's grey literature.

- ✓ The public version of the DOE Information Bridge (<http://www.doe.gov/bridge>) offered in collaboration with the U.S. Government Printing Office (GPO) Depository Library system, has over 43,000 DOE R&D grey literature reports (more than 2.8 million pages) that are electronically available and free to the public worldwide. OSTI is proud to have been among the first sites to make such a large volume of what used to be considered grey literature available to the Internet literate, winning several governmental awards for its efforts. Over the past 6 months, the two Information Bridge products combined averaged more than 2200 full text reports being downloaded weekly, far exceeding traditional distribution methods of the past. The grey literature is not simply available, but it is being USED.

The paradigm of grey literature being hard to get has indeed shifted for energy-related STI. So, how did we get there?

#### FIVE YEARS AGO

OSTI has always had responsibility for DOE's grey literature. For many years DOE and its contractors were required to provide paper copies of their reports to OSTI for public dissemination. Storage and dissemination at OSTI changed over time from paper, to microcards, and then on to microfiche. A bibliographic database citing the reports and other commercially published energy information was also created. Citations included abstracts and added subject indexing to facilitate retrieval of the information. Public dissemination of the DOE grey literature was accomplished through partnerships with the GPO Depository Library system and with the Department of Commerce's National Technical Information System (NTIS). Microfiche and paper continued to be the only ways to get this grey literature for many years. The constant challenge was to improve the timeliness of dissemination, because of the resource intensive nature of the storage and retrieval methods available to fill orders.



In addition to collecting and disseminating its own DOE grey literature, OSTI has a long history of international exchange of grey literature and related bibliographic information. OSTI's chief multilateral agreements are with the International Energy Agency Energy Technology Data Exchange (IEA/ETDE) and the International Atomic Energy Agency International Nuclear Information System (IAEA/INIS). ETDE, begun in 1987, currently has 18 countries exchanging the full scope of energy-related information (<http://www.etde.org>) while INIS, begun in 1970, has over 100 countries and 19 international organizations exchanging primarily nuclear-focused information (<http://www.iaea.or.at/inis/inis.htm>). From the inception of these agreements, OSTI has served as the US delegate and met US commitments as well as possible given recurring budget restraints. Since ETDE's inception, OSTI has managed and operated ETDE on behalf of its members as Operating Agent. As within DOE, five years ago the paradigm for grey literature exchange in these agreements was also paper and microfiche.

#### DRIVERS FOR CHANGE.

The 1990's of course brought better and faster computers, much wider use of the Internet, and a vastly more computer literate workforce. At the same time, reduced budgets and downsizing of government became the norm. OSTI in particular saw its budget decline by nearly half even though information and technology were continually recognized and touted as the new wave of the future. While new technologies are allowing for cost efficiencies, an up front investment is usually required. How to cope in this new environment with the new resource limitations became a formidable challenge. User expectations began to shift rapidly. Few would argue that the Internet considerably influenced the mind set of users who now want instant gratification. Information users today have little patience to wait weeks for a copy of a document; even waiting minutes for file downloads now seems like an eternity. The concept that because something is free it has less value has decidedly lost its influence. Many users today gravitate to what's easiest to get rather than the best or most scientifically correct. Bibliographic information alone is passe – users demand full text... Now!

To address resource issues as well as the new environment, OSTI knew it had to take advantage of new technologies and work with its partners wisely. It was no secret that more and more documents were being created in electronically shareable formats – as opposed to the sometimes cumbersome strict publishing formats of the past. Sites throughout the DOE community were evolving internally to new ways of producing their reports. Technology offered TIFFs, GIFs and PDFs as new options in the electronic viewing and printing of documents, without requiring users to have specific word processing software. Full text



searching was becoming a real possibility, bringing with it some schools of thought that bibliographic information would someday be obsolete. The vision, as OSTI saw it, was to give information to the user at the desktop, and this meant electronic full text.

In 1996, OSTI made a strategic decision to migrate to electronic full text and transition out of microfiche production. The challenge: what format(s) do you standardize on? Since DOE documents were continuing to come to OSTI in paper form, the format chosen for storage and manipulation was TIFF Group IV, and a scanning process was established. Microfiche production continued in parallel for some time, but was modified to use electronic TIFF images instead of the traditional paper and camera methods.

The next challenge was to choose a dissemination outlet. Thus the DOE Information Bridge concept was born. This Web concept had the bibliographic record retained as the base for searching but added a hyperlink to the full text. In addition, some full text searching capability was built into the product line. This was achieved by an automated OCR process of the scanned TIFF images to build a text index. Although no manual clean-up of the data was done, a good quality copy would produce quite good results, and served to enhance bibliographic searching.

Another challenge faced during this time frame was whether access should be free. Should users see the information for free? Should they be able to download whole documents or only single pages? As a government entity, OSTI's work was already funded by taxpayer dollars. But this was added-value, and extra service. Then again billing meant accounting systems and additional resources. The end result was free access, and the capability for both page at a time viewing and full document downloads.

The next goal was to get partners involved. In collaboration with the DOE STI partners, regulations were revised to state that the preferred method of receipt of DOE reports would be electronic, and that soon dissemination too would be electronic. In early phases, only 5 formats were to be permitted on receipt: SGML, HTML, TIFF Group IV, PDF, and Postscript. The receipt format became known as the native format. At the time, SGML looked the most promising for scientific exchange of information, although experience has not borne that out. While OSTI does receive a few documents in SGML, the largest majority have been in PDF, followed by TIFF, then HTML, along with still quite a lot of paper. In recent times, OSTI has added acceptance of some word processing formats, but receipts to date have been few in number. TIFF images continue to be created from any paper received.

Regardless of the native format, OSTI initially chose TIFF Group IV as the standard for storage, dissemination and archival purposes. Over time, however, PDF has taken a dominant role, and PDF-wrapped TIFF images are also offered for dissemination for all documents residing on the Information Bridge. It is not clear yet whether OSTI will move to only storing PDF for archive, but it is a real possibility given the storage requirements to handle native, PDF and TIFF copies of each document. The challenge of bulk delivery of electronic full text was met by choosing the 8mm DAT as the output media for the TIFF images (in addition to their availability via the Information Bridge product line.) Short SGML files for each document accompany the TIFF images on DAT, to provide brief bibliographic information.

The next challenge was to involve international partners. OSTI's move to electronic full text certainly had impacts on these partnerships. OSTI worked closely with the INIS Secretariat in Vienna, Austria to share future directions and ideas, since US information made up a considerable portion of the INIS system. INIS also chose to migrate to electronic full text, with TIFF as the primary storage format following their own scanning procedures. However, INIS chose a CD-ROM companion product to the INIS bibliographic database as their dissemination media to members. With a number of countries in INIS still having limited Internet access, the CD-ROM option was chosen as the best choice to make full text available to all member countries. The main drawback of this system has been the number of CD-ROMs necessary to hold the data. INIS is currently looking at DVD options to condense the number required for a full collection, but much depends on the availability of the technology in those same countries where Internet is not a universal solution.

INIS has also chosen to transition out of microfiche dissemination, but decided to retain microfiche production for archival purposes, although it too is produced from TIFF images. In addition, OSTI receives an 8mm DAT of TIFF images from INIS to load as part of the limited access Information Bridge product, and provides INIS with the US full text information on DAT.

With OSTI as the ETDE Operating Agent, plans were shared within the ETDE community. ETDE member countries were encouraged to submit data electronically, and the Information Bridge and 8mm DAT dissemination were offered to ETDE input centers. Several countries receive the electronic full text to serve users in their country more directly. As Operating Agent, OSTI scans in grey literature in paper form received from ETDE members to create electronic full text. This is then added to the Information Bridge limited access product. While some documents are received at INIS and ETDE in electronic format, the large majority



continue to come in paper form. The last year has shown movement in some countries, however, to improve this situation.

Since 1996, many issues have arisen and have been met regarding electronic full text. Some that continue to be considered include which formats should be allowed, and which should be considered the archive. How can searching be improved? An issue related to ownership arose within the international community, that DOE was fortunate enough not to face. Some full text sources who were perfectly content with allowing microfiche copies of their documents to be distributed were not so eager to allow the proliferation of electronic versions. For some countries, this caused the unexpected side effect of a decline in the number of documents that could be made available by ETDE and INIS, making that literature even more grey in terms of ready access.

As electronic availability became more the norm within the DOE community, OSTI has offered DOE sites the option to send a URL to their documents, if the sites already have them on Internet. Within the Information Bridge structure, users can still hyperlink to the documents, but they will not physically reside at OSTI. Thus, storage of electronic full text is moving to a decentralized model even though retrieval continues to be facilitated by a centralized database. Archive issues remain a concern with this model, but sites are supposed to notify OSTI before eliminating their Internet access. Coming soon will be OSTI's implementation of persistent URLs or (PURLs) to allow more direct access to documents and to minimize the impact on users if documents are moved to a new location.

### STILL TO COME?

For OSTI, new versions of the Information Bridge product line will soon be available, adding some enhancements for users. A popular request from users that will not be implemented right away due to cost is to make the full text searchable once downloaded. Neither TIFF nor the PDF-wrapped TIFF allow such functionality. While OSTI has looked into various options for doing so, all take time and resources.

EnergyFiles (<http://www.osti.gov/EnergyFiles>), OSTI's virtual library product that directs users to all types of energy information, continues to grow, with distributed searching options now being offered. In addition, a new product called PubSCIENCE (<http://www.osti.gov/pubsci>) debuts in October and will provide users with the capability to



search across a large compendium of peer reviewed Journal literature with a focus on Physical Sciences and Technology. Online access to the full text of many journal articles has been a much desired next step expressed by researchers. Although viewing of the full text from most publishers is not without cost, PubSCIENCE centralizes access to many journal articles without having to leave the office. In some ways, while journal literature has never been considered grey literature in the traditional definition, almost any paper-only product line could be considered grey literature in today's electronic environment.

Phys Sci  
Tech

Change and  
def?  
? going

In the international arena, ETDE has opted to have an Information Bridge-like product available to users in their countries beginning in late October. The product, to be called ETDEWEB or ETDE World Energy Base, will be offered from the ETDE Web Site at <http://www.etde.org>.

OSTI feels proud to have done it's part to meet the challenge of removing the "grey" from grey literature in energy by making it accessible. It has done so despite limited resources and with some difficult decision-making along the way. Without the commitment of a dedicated and capable staff and willing partners, grey would still be grey, and OSTI would be going down the path of the dinosaurs.

Contact information:

Debbie Cutler

International Programs Manager & IEA ETDE Operating Agent Representative  
DOE/OSTI

175 Oak Ridge Turnpike

Oak Ridge, TN 37830

Ph: 423-576-1272

Fax: 423-576-2865

e-mail: [debbie.cutler@ccmail.osti.gov](mailto:debbie.cutler@ccmail.osti.gov)

LINK MANAGERS FOR GREY LITERATURE

Elena Lodi, Dipartimento di Matematica di Siena

Martin Vesely and Jens Vigen, CERN

Abstract

In the self-service area of the library reading rooms it is necessary to organise the collections in the simplest way possible. This is an important feature for readers, making it possible to get direct access to the material without necessarily having to go via the library catalogue to retrieve the call number. A typical example will be that a collection is organised so that a reader can easily get directly from an article reference to the article itself in a library where the journal collection is simply organised alphabetically by title.

This requirement seems to have been forgotten by many of the most important actors in the digital library field, both by the commercial publishers and by many of the bodies producing grey literature. How are the users supposed to get directly to report xyz without having to navigate through n different web pages? This is happening as well in "The Library" which is intended to be close to 100% based on self service!

CERN Library has developed a mechanism called "Go Direct", to handle this problem. So far it only processes articles published in journals, but due to its success it will however be extended to include grey literature issued in series holding a report number.

The CERN Library catalogue uses "Go Direct" to automatically generate links from bibliographic information to the corresponding fulltext of a document.

The talk gives an overview of the present situation concerning link managers and URL architecture in the field of electronic publishing within high-energy physics.

### Introduction

The CERN Scientific Information Service has for more than forty years been collecting and organising information of interest for the whole high energy-physics community. The aim is to have a collection as complete and up to date as possible. With the explosion of the Internet and the introduction of preprint servers the amount of information has multiplied several times over the last ten years, in parallel with the fact that user requirements are much higher today than ever before. Our Service is now processing around thirty thousand new grey research documents per year in addition to "whitewashing" thousands of other documents collected in earlier years.

With the introduction of electronic preprints (Dallman et al., 1992) followed by electronic journals (Chaney et al., 1999a), completely new possibilities for weaving the information together have arisen and so have new problems. In the pre-electronic era it was possible to get directly from a reference to the corresponding document in the library, often without using a catalogue, while now the same operation would require a significant amount of navigation through a number of web pages. The CERN



Scientific Information Service's main aim in organising the information has been to avoid unnecessary navigation and to introduce new identification systems. Related documents are linked together and the system permits the users to get to a given document simply based on the bibliographic reference without having to go via hierarchical trees or any other "clicking exercises".

The very first solution to this problem was to have a simple form with a script behind it which, based on the submitted information, could compute the URL of the required document. The idea was to provide a facilitating tool for the scientists so that it should be easy to retrieve the articles available online. As the form, now called "Go Direct", became successful, it also became clear that the script could instead take its arguments directly from the library catalogue so that links could be created "on the fly".

#### Collection and manipulation of information

Most of the acquired documents in the CERN Library which are considered as grey literature originate from the Los Alamos Eprint Archive, but there is also a substantial amount of documents originating from inhouse CERN research and direct submission from other institutes. The collected information, the metadata, is unified and presented within WebLib, the web interface to the library catalogue and various other CERN databases. This interface enables easy access to the metadata and other features such as ordering documents, loan renewals and access to the fulltext for a significant amount of the documents.

As soon as a document is introduced into the system, it is automatically analysed. The references are extracted from the file and stored separately in order to allow later citation searches. The extracted results are then processed through a "normalisation filter" in order to ensure that the linking program will recognise the references. This means that correctly analysed references, referring to online available documents, will actually point the users directly to the corresponding fulltext.

Most scientists prefer to read the published version of an article when it is available, instead of the preprint. Consequently an important role of a library handling masses of grey literature is to connect such documents together as soon as the published versions appear. This procedure in the CERN Library is handled fairly automatically for documents published in the journal literature by a program that queries WebLib with the intention of matching and updating the preprint records for the documents which were published.

For each article, depending on the source of information, different metadata will be available. Normally at least the three first authors will be supplied, the year of publication and the published title in addition to the publication reference itself. The database index is such that the answer to a query is faster if the search is done on the basis of consecutive words. For this reason the program selects the longest phrase of consecutive words from the title. A word must contain only letters. Moreover some words, like the logical operators *and*, *or* and *not*, are stop words. A phrase is a consecutive sequence of words separated by spaces. For example in the title *Hadronic ZZ, W W', and production with QCD corrections and leptonic decays*, there are 3

phrases: namely *Hadronic ZZ production with QCD corrections* and *leptonic decays* of length 2, 4 and 2 words respectively. The program will then choose the phrase *production with QCD corrections* as an argument to be supplied to the matching algorithm.

A query submitted to WebLib cannot be longer than 74 characters and cannot contain more than 7 logical operators. The program makes the query using all the available authors, the longest phrase of words in the title and the year 1997, which covers all the years in the range [1990,1999]. If the answer to the query returns one or more records, the titles are examined by a pattern matching algorithm. The problem of pattern matching is the following: given two strings called a text  $T$  and a pattern  $P$ , if  $P$  occurs in  $T$ ,  $P$  is a substring of  $T$ . Let  $T$  equal to  $\langle X_1, X_2, \dots, X_m \rangle$  and  $P$  equal to  $\langle Y_1, Y_2, \dots, Y_n \rangle$ , where  $m \geq n$ .  $P$  will then be a substring of  $T$  if there exists a value  $i$ ,  $1 \leq i \leq m-n+1$ . Consequently the string  $\langle X_i, X_{i+1}, \dots, X_{i+n-1} \rangle$  is equal to  $P$  when  $X_i=Y_1, X_{i+1}=Y_2, \dots, X_{i+n-1}=Y_n$ . For  $T=\text{Parallel updating cellular automaton models of FK systems}$  and  $P=\text{cellular automaton models}$ ,  $P$  is a substring of  $T$  starting from position  $i=19$ . If  $P$  is equal to  $T$  the problem is called perfect matching (Crochemore and Rytter, 1994).

Approximated pattern matching (Bertossi et al., 1990) is an extension of the pattern matching problem which allows for the possibility of having errors in the matching. A posteriori the most common errors of typing are the following:



- a mismatch, in the sense that the two corresponding characters in  $P$  and  $T$  are different
- a missed character in  $T$
- a missed character in  $P$ , that is a character more in  $T$ ,
- inversion of two consecutive characters.

For example let  $T = \text{Non-universal Quasi-Long Range Order}$  and  $P = \text{Nonuniversal, Quasi-Long-Range Ordre}$ ,  $P$  is recognized as a matching of  $T$  with 4 errors: a missed character in  $P$  ( $X_4 = "-"$ ), a missed character in  $T$  ( $Y_{12} = ", "$ ) a mismatch ( $X_{22} = " "$  different from  $Y_{25} = "-"$ ) and an inversion ( $X_{35} = Y_{36} = e$  and  $X_{36} = Y_{35} = r$ ).

In the program there is also a treatment of formulae. In WebLib formulae are given following the syntax of LaTeX which means that they will be delimited by \$ signs. Superscripts and subscripts are indicated by ^ and \_ respectively and their values are within parentheses, usually | } Many of the information sources used for the updating eliminate all these special symbols from the titles, consequently the matching program is also using the "linear" version of the title too. For example the title = *Search for  $\eta_c'$  and  $h_c$  ( $^1P_1$ ) states in the  $e^+e^-$  annihilations*, will appear in latex as *Search for  $\eta_c'$  and  $h_c$  ( $^1P_1$ ) states in the  $e^+e^-$  annihilations*, becomes typically *Search for  $\eta_{c'}$  and  $h_c$  (1P1) states in the  $e^+e^-$  annihilations*.

When a preprint is accepted for publication it will normally be revised when going through the referee process and the title itself might even change. The most frequent changes in the title are added words. The program finds matches on the basis of single words, takes as pattern the shortest title in words and allows missing words. For example *From the Hubbard to the SO(5) ladder: a numerical study* is recognized as a substring of *Evolution of the low-energy excitation spectrum from the pure Hubbard ladder to the SO(5) ladder: a numerical study*.

Finding matches on single words makes it possible to find abbreviations as the following example shows: *Six-body LFTD and wave functions for the massive Schwinger model* is recognized in the title *Six-body Light-Front Tamm-Dancoff approximation and wave functions for the massive Schwinger model*. The program counts errors on the basis of missing words. For example, if the title furnished by the information source is *Chiral odd structure functions from a chiral soliton* we can find the following three titles in WebLib:

- Nucleon structure functions from a chiral soliton
- Structure functions from chiral soliton models
- Chiral odd structure functions within a chiral soliton

All these three titles are at distance 1 from the given text. In fact when the program analyses the first title, it finds all the words in the pattern except the first one, *Nucleon*. In the second case the missing word is *models* and in the third *within*. A

human being can decide quickly that the third title is the one to choose, but the algorithm can not. For this reason the database is automatically updated with the references to the journal, volume, issue and page only when so called perfect matches are found. In the other cases the approximate matches are written into a file that will be further analysed by a librarian. The work of the librarian is facilitated by an additional program that displays the bibliographic data of the article on the right side of the screen, and on the left side the data of the matching candidates. The program also provides links to the fulltext so that the librarian, if necessary, can compare the actual texts.

The matching program is written in Perl and has more than 4000 lines of code. It takes only one parameter, a list of published papers or simply a table of contents, (let's call it INDATA) and produces 4 files called `verify.indata`, `noconfidence.indata`, `correct.indata`, `append.indata`. The first and second files are created in the same directory where the program is located. The file `verify.inadata` is a file of verification to check if errors occur during the processing of the data. The file `noconfidence.indata` is the one used for the treatment of approximate pattern matching. The files `append.indata` and `correct.indata` are later run in batch mode to update the catalogue.

#### Automatically linking to the fulltext

Traditionally a preprint was simply retrieved in the library collection by its report number in the same way that an article was retrieved by its triplet journal title, volume and page. Moving into the electronic age this was paradoxically not as simple



is part ✱  
anymore. In most cases the users were obliged to point their browsers to different preprint archives or publishers sites and start navigating from there, even if the reference of the document they were looking for was already known. CERN scientists wanted, knowing the reference, to get directly to the fulltext of the documents. The whole project started in 1997, not a long time ago but still in the early days of electronic journals in the CERN Library (Chaney et al., 1999b). Looking at electronic journals, at that time only Physical Review D, published by the American Physical Society (APS), had chosen to publish using an URL architecture making it possible to associate a bibliographic record with a fulltext article without having to store the URL in the database (Doyle, 1999). The term "link manager" hardly did not exist. Since then a long period of lobbying other publishers to follow the example of APS has followed. Surprisingly enough this has not been straightforward, the approach has apparently been too pedestrian for many of the important players in the field. Today is it however adopted by the Institute of Physics (IOP), EDP Sciences (EDSP) and Elsevier has implemented it for a few important titles in the field of high-energy physics. In addition there are publishers which can be handled by maintaining a set of look-up tables at the CERN servers.

Having thousands of publication references in the library database it became clearly interesting to automatically locate the online published articles using Internet technologies, especially as articles were also made available retrospectively, predating the electronic preprint archives.

The electronic journals are organized in remote databases within the publishers information systems and are accessible via the Internet restricted to the subscribers. This implies the need to introduce some Internet-based mechanism that would provide the connection between the bibliographic information available from the library catalogue and fulltext of the corresponding document.

Such a tool has been developed, aiming to serve the WebLib which can be accessed and browsed from distant terminals and search a variety of databases independently of their location. This combination became a very comfortable tool for document research. However, since no identification standard for electronically issued documents has been established, the online fulltext version of published articles could not be located in a trivial way.

Recently the topics of identification of electronic documents have been discussed widely. The need for a unique identifier lead to several de facto standards for identification, i.e.: Digital Object Identifier (DOI), Publisher Item Identifier (PII) or Serial Item and Contribution Identifier (SICI) (Doyle, 1999). For the location of a document on the Internet the given reference needs to be transformed into a URL through which the document is accessible. The URL is in fact not a permanent identifier and some standardized way of referring to online documents should be established. However, such a standard has not yet been accepted by a majority of the community. In addition to that, the well established system for referring to printed material, using the triplet journal title, volume and page, is still around and will



continue to be for many years. It appears therefore natural for the time being, to build an identifier based on the traditional standard reference.

Before the link manager was implemented, the most obvious solution was to store the particular link in the library catalogue and whenever the full text was required, the URL was available directly from the catalogue. This way the catalogue records contained both the standard reference and the URL itself. By introducing another field in the bibliographic record, this solution was easy and straightforward. On the other hand, the amount of articles that had to be handled required too much manual maintenance. Additionally URLs are not stable and can be freely modified at any time. The update of the modified URLs would be even more demanding, even if logically possible, because it would require periodical checks throughout the entire database. The acquisition of metadata could have been automated by building an engine that supplies the bibliographic records to the catalogue with a unique URL for each article. Eventually, the development of this engine would be much more powerful embracing a different approach to the solution.

The idea is to not store the URLs in the catalogue at all, based on the fact that the URLs are not considered as identifiers and therefore they might be variable. The valid URLs are then composed at the very moment the requests arise. In order to be able to perform this composition the system includes another independent module: the link manager. The link manager is a service based on web technology that provides applications with valid links to specified documents online. The idea is to keep the



links up to date by moving the responsibility for their validity from static storage in the catalogue database to a resolution mechanism. Unlike the database, this resolution mechanism does not store any links, but creates them dynamically according to the current configuration. In case the final URL changes during its lifetime, the link manager has only to be reconfigured. The link that appears in the client application is then replaced with the address of the link manager that composes the final URL from the given reference.

The mechanism consists of two parts that cooperate as necessary (link builder/indexing robot), where one takes care of solving the requests and the other one maintains the metadata repository. The metadata repository contains firstly the configuration that defines how the URL will be designated in particular cases – for certain documents or sources. This configuration is the fixed and required input for the link manager to work. Secondly, in case of handling non-standard references, the system needs to index some additional information, which can be gathered using the indexing robot. This indexing has to be performed periodically as a batch process, usually triggered by date. [The link builder then has three inputs: the reference asked by the client, the configuration for the particular source and the indexed information.]

The link manager module can be used by any application that has to deal with linking to fulltext documents. As mentioned before, besides WebLib the link manager is used for the form "Go Direct" that allows the user to request articles without going via the catalogue.

In the WebLib solution (fig. 1) the link manager provides the connection between the bibliographic information available from the library catalogue (local database) and the fulltext information provided by the publishers over the Internet (external database). The range of Internet searched locations is significantly restricted. The robot goes only through those locations that are defined in the metadata repository as bibliographic information resources - in this case the subscribed electronic journals.

The link manager resolves the standard reference passed through WebLib and returns the valid URL. No URLs are stored in the library catalogue or in the metadata repository, but are created "on the fly". When invoked from the web application the link manager automatically switches to this determined location so the document can be retrieved.

The link manager has to be flexible enough to be able to link to many different types of locations. The system retrieves data from different publishers and all of them support different strategies for data storage and therefore different URL structures. The download of metadata and its indexing has only to be performed when the actual URL contains information that cannot be derived from the standard reference. In many cases this can be the issue number within a certain volume, the internal document number assigned by the publishing house, your client identification or any other information that for some reason or another can be used as an identifier.

A few publishers have their own link manager service, which simplifies the URL designation on the client side, because the system only has to deal with the standard supplied by the publisher. In this way is it possible to build some generic algorithms dealing with the particular standards. In all other non-standard approaches special algorithms have to be defined.

The URL then contains the location of the particular link manager supplied by the publisher and the standard reference. Looking at an example, the APS link manager (<http://publish.aps.org/linkfaq.html>) locates the article in the following way:

[http://publish.aps.org/abstract/\\$journal/v\\$volume/p\\$fid](http://publish.aps.org/abstract/$journal/v$volume/p$fid)

where the following variables must be supplied:

\$journal	abbreviated journal title
\$volume	volume
\$id	starting page

The URL for any APS article, generally presented, will then look like:

<http://<link manager address>/<formatted standard reference>>

which then passes the request to the external resolution machine.



Similar for the non-standard references:

`http://<server address>/<directory location>/<fulltext file>`

which links directly to the article.

In the final state the system provides the connection using two resolution machines, one on the client side and one on the publishers side (fig. 2). The local link manager allows the unification of the records stored in the library catalogue, enabling access to many different sources. It is also desirable to choose the identifier independently of the external environment. In this way is it possible to use the standard reference (the triplet) also as an identifier for online versions of the electronic documents, without introducing other fields in the catalogue. The local link manager then transforms the defined identifier to the standard that so far depends on the agreement between the two communicating sides. The remote link manager will eventually resolve the final location of a document that is being retrieved. The "two link manager solution" simplifies the process of linking to the fulltext of any article by defining the interface on the application layer.

Systems with a link manager only on one side can also be implemented. The local system will then have to take over all the work and it therefore becomes significantly more complex. Such a solution requires building up a quite large metadata repository

in order to also be able to handle the non-standard references which of course is more difficult to maintain. If the link manager is always available remotely, the local link manager is in fact outsourced and to switch between different sources, a simple CGI script can be then used instead.

At CERN some documents are provided directly by the local information system. Therefore the system in fact enables quite flexible two-way exchange of scientific information by both linking to external resources and allowing other institutes to link directly to local fulltext documents.

### Future challenges

The developments which were meant to make the world simpler for the scientists made it in some senses actually more complex. The documents are in principle easier to get access to, but there are more references to a single document and its published counterpart than there were ever before: author/title, report no., preprint archive no., URL, DOI etc., etc. and finally the traditional bibliographic citation - or at least it is still there for the time being. The goal must be to reduce these different references to a minimum and establish a standard which is as intuitive as possible. Having made the major publishers aware of the power of the link managers and later getting several of them onboard, the situation seems to be turning in the right direction. The work of lobbying the publishers, also the ones issuing grey literature, is however not yet at an end. To augment the pressure the campaign should also be directed to other database

providers. The more people applying this simple system, the more likely it will be that it will become a de facto standard. The major challenge for the future will be however to find a better solution for how to associate grey literature with its published counterparts. The DOI Foundation, a foundation more and more publishers are joining, is just aiming for this. The DOI is meant to be a unique identifier of any piece of intellectual content, together with a system for using that identifier to locate digital services on the Internet associated with that content (Paskin, 1999). In the field of high-energy physics there is already a veritable preprint culture which is accordingly well organised. More or less all these preprints have report numbers and it is not difficult to imagine that they accordingly could be assigned DOIs. Exploiting the power of the DOI system, one could then reach a situation where by pointing to the DOI of a preprint, the researcher would automatically be offered the option to follow a link to the published counterpart as soon as this would be available.

#### References:

- Bertossi, A. A., Lodi, E., Luccio, F., Pagli, L. (1990) "Two string matching problem with weighted errors", *Theoretical Computer Science*, Vol 73, pp. 319-328
- Chaney, E., Bulliard, C., Christiansen, C., Cressent, J. P. (1999a) "Une bibliothèque de recherche face à l'édition électronique: l'exemple du CERN", *Bulletin des bibliothèques de France*, Vol 44 No 2, pp. 27-32



Chaney, E., Bulliard, C., Christiansen, C. (1999b) "The electronic journal service at CERN, a first evaluation: user access interfaces and user awareness", *Vine*, No. 110, pp. 23-29

Crochemore, M., Rytter, W (1994) *Text algorithms*, Oxford University Press, New York

Dallman, D., Draper, M., Schwarz, S (1994) "Electronic pre-publishing for worldwide access: the case of high energy physics", *Interlending & Document Supply*, Vol 22 No 2, pp. 3-7

Doyle, M (1999) "Citing and linking in electronic scholarly publishing: a pragmatic approach", edited by Smith, J. W. T., Ardö, A., Linde, P., *Electronic Publishing '99. Redefining the information chain – new ways and voices: proceedings of an ICC/IFIP conference, Karlskrona/Ronneby, Sweden, 10-12 May 1999*, ICC Press, Washington D.C., pp. 51-59

Paskin, N. (1999) "The digital object identifier system: digital technology meets content management", *Interlending & Document Supply*, Vol 27 No 1, pp. 13-16

Figure 1

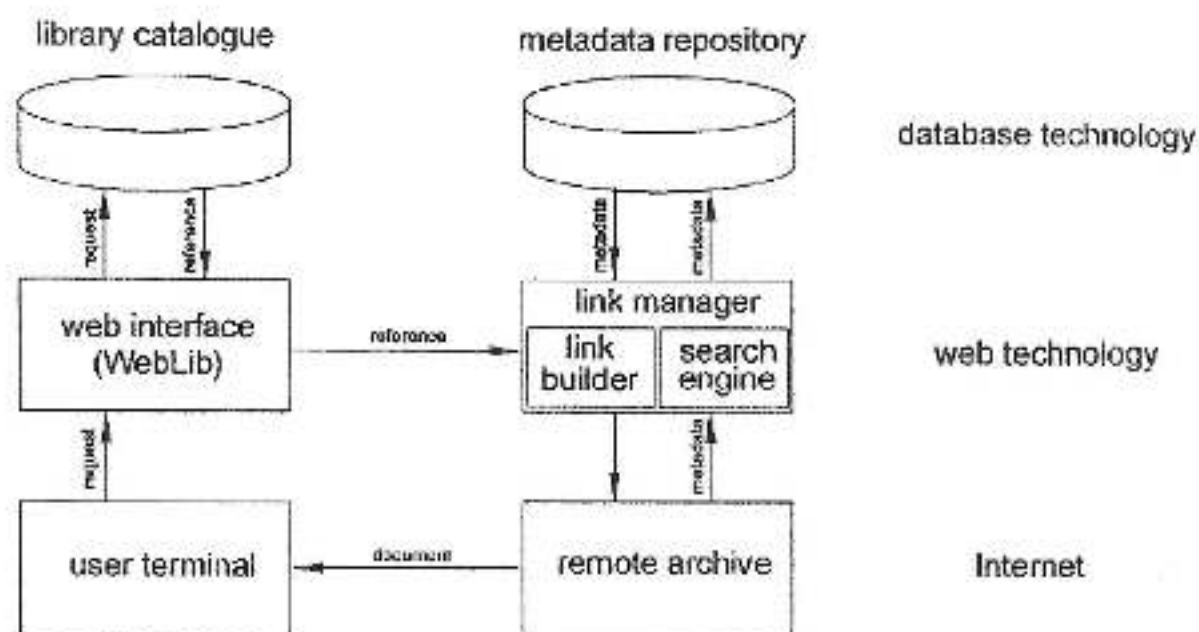
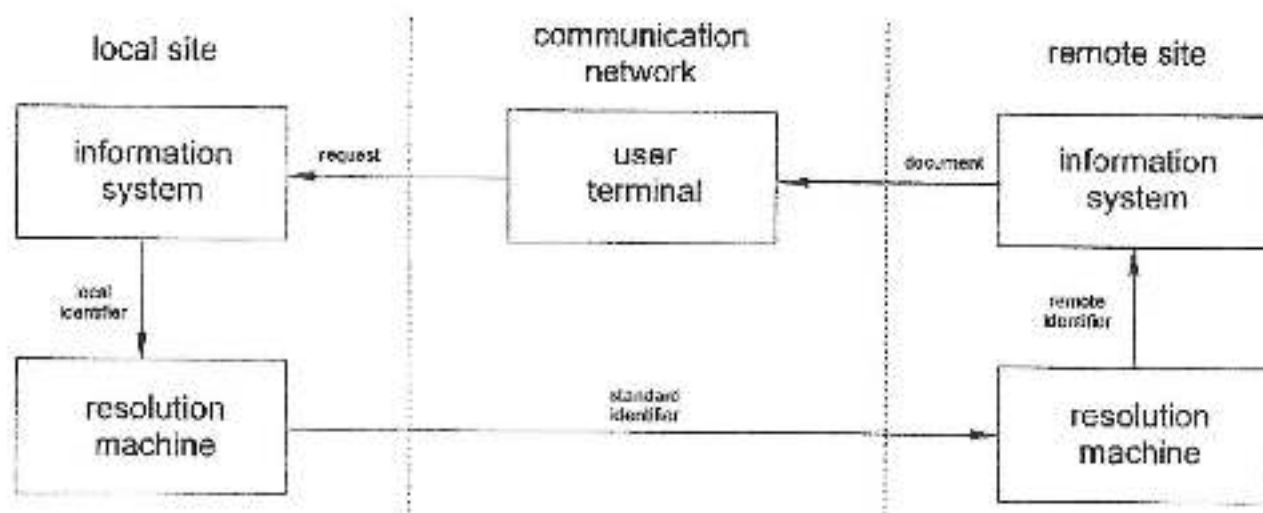


Figure 2



## Grey Copyrights for Grey Literature:

### National Assumptions, International Rights

#### AUTHOR

Michael Seadle  
Digital Services and Copyright Librarian  
Michigan State University  
100 Library  
East Lansing, MI USA 48823-3533  
Editor, *Library Hi Tech*

#### ABSTRACT

Copyright enforcement is a mixture of law and culturally-influenced practice. Although the statute law appears in black and white, the traditions and expectations of those who create and use gray literature are definitely grayscale. And the latter are a better predictor of both infringement and litigation than the letter of the law itself. Casual US assumptions about using gray literature stem in part from technicalities in the pre-1978 copyright law, which often threw gray literature immediately into the public domain. These assumptions stem also partly from the low economic valuation of gray literature, and the key role economic value plays in US copyright litigation. Since moral rights plays almost no role in US copyright law, common European features such as the right of withdrawal have no place in local expectations. In recent years digital publishing and international access have exposed these assumptions to a context where the separation of moral and economic rights is taken for granted. Legal rights, both within the US and Europe, have also come in conflict with the urgent desire of both scholars and librarians to make gray documents more accessible by republishing them on the Web.

[ARTICLE BEGINS]

#### INTRODUCTION

"This, no doubt, is why natural history, in the Classical period, cannot be established as biology. Up to the end of the eighteenth century, in fact, life does not exist: only living beings." -- Foucault, 1973, p. 160

It would be easy to paraphrase Foucault and say: until the end of the 1960s, grey literature did not exist, only uncontrolled documents. Grey literature as a concept is relatively new. Everyone realized that some documents did not get into standard bibliographic tools, did not appear in library catalogs, did not, in fact, have the status of "real" publications from legitimate commercial presses. An archivist might squirrel them away in Hollinger boxes with a cryptic label and a note on a typed finding-aid that probably listed them under the name of the person or organization who was peculiar enough to collect them. They were the untouchables of the library world, a permanent low caste because they served purposes tainted by self-promotion, private



interest or practical application. They had, in the most literal sense, little or no value. They were not even bookstore outcasts, since they never once reached the shelves.

Copyright issues for grey literature need to be understood in this human and historical context. The law itself is no different for a mimeographed pamphlet run off in blue ink by a volunteer organization than for a hardcover book with crisp, stylish printing, and a professionally designed dust-jacket. But the attitudes toward these different works, both for creators and consumers, diverge because of the relative value assigned to each. When the physical object appears to be worth more, the intellectual property tends to be treated accordingly.

Property rights depend in part on the willingness of people to defend them. The owner of a piece of land who allows public passage across it may in time lose the right to sue for trespass, perhaps even to deny access. The same is true for intellectual property rights, especially in the United States, where, until 1989, notice and registration requirements were an integral and important part of the law. The author of a work of grey literature had to care enough about it to comply with the notice requirements to have any hope for protection. Many did not. In fact, many intentionally did not because they saw copyright only as a restriction on the dissemination of information that was their real object.

The tensions between law and practice are illustrated below in three examples. In the first, the creators of the grey literature deliberately ignored copyright requirements in order to improve access to their message. In the second, the rights holders claim has come under attack because of initial irregularities, and their rights have been hard to enforce. In the third, copyright protection was unambiguous and was actively claimed, but open, blatant infringement still occurred. To understand these examples, a brief overview of US copyright law may be helpful.

### AN OVERVIEW OF US COPYRIGHT LAW

This overview covers five issues relevant to grey literature: 1) notice and registration, 2) duration of protection, 3) published versus unpublished works, 4) moral rights, and 5) enforcement.

No one should read this overview as legal advice, or even as an accurate portrayal of the full range of complications of copyright law in the US. Those who wish to examine the statutes themselves can find them on the Web courtesy of Cornell University Law School. [1] Those who wish to learn some of the basic case law may wish to consult a compilation like West Publishing's *Cases and Materials on Copyright...* (Nimmer, 1991)

#### Notice and Registration

The notice requirement went through three historical phases. The first lasted until the major 1978 revision of the copyright law. In it, notice in the form of the word "copyright," the abbreviation "copyr.," or the symbol "©" was a prerequisite for protection. Omit it, however accidentally or unintentionally, and the work, if publicly distributed, fell immediately and irretrievably into the public domain. The second phase lasted from 1978 until 1989. The law still required a notice for published works, but allowance was made for minor accidental omissions, which could be rectified without permanently losing protection. The final phase runs from 1989 to the present. Under it, the notice requirement has been abolished in order to bring US law into accord with the Berne Convention. Of course notice applied only to text. The status of spoken words, broadcasts, and performances, though generally protected, was more ambiguous until 1978.



The registration requirement, like the notice requirement, became less absolute after 1978, but remains necessary for some types of damages. Post-infringement registration is also possible. There is a \$30 registration fee – low enough not to be impossible for the poor, though perhaps still high enough to daunt the frugal.

#### Duration of Protection

Before 1978, the US law protected works for a maximum of two 28 year terms. The second term was not automatic until 1964, which meant that many works received only the first 28 years of protection. In any case the absolute maximum was 56 years. After 1978, the term of protection followed the Berne Convention standard of the life of the author plus 50 years, or 75 years from publication for anonymous works. This meant that accurate knowledge of the death date of the author became more important in determining the term of protection than the date of publication, which could generally be found even in a grey literature work. In 1998 the period of protection increased to life plus 70, or 95 years for corporate or anonymous works.

#### Published vs Unpublished Works

Before 1978, unpublished works did not come under the federal copyright statute at all. They were governed instead either by state or by common law. The latter gave unpublished works perpetual protection, with no need for a copyright notice, no need for registration, no need in fact for anything except that the works be fixed in a relatively permanent form. This meant that letters, personal papers, and any other works which never met the definition of publication had enduring protection. Works which had a limited distribution among a private audience, such as the draft of an article or book, could well (and generally did) fall into this category.

After the new copyright law came into effect in 1978, unpublished as well as published works received protection from Federal law. A grace period prevented unpublished works from falling into the public domain until the end of 2002 at the earliest (if created before 1978) or 2047 (if created after). Except for these limits, the usual rule of life-of-the-author plus 70 years now applies to all published and unpublished works (95 years for corporate and anonymous authors).

#### Moral Rights

Moral rights for text or phonorecords have no place in US law. The US Congress did enact limited moral rights protection for the creators of works of visual art as part of the bargain for joining the Berne Convention. But none of the extensive set of European rights that allow an author to prevent defacement or to withdraw a written work exist in any form in the US statutes. Like the Anglo-American tradition generally, the US copyright law focuses almost exclusively on economic rights.

In Germany and France, moral rights are inalienable. The creator cannot sell them or give them away, though they can be inherited. They are also personal rights, and a corporation cannot own them. No such restriction exists in US law. In fact, the trend toward corporate ownership of copyrights seems to be growing.

#### Enforcement

Rights owners must choose whether to enforce their privileges or not. It is neither cheap nor easy to enforce the copyright law against a determined infringer. A private letter to an infringer may work, but most rights owners start with a letter from a lawyer. If the infringement continues, pursuing the issue means going to court. It also means having to register the work, pay the fee, deposit at least one copy of the work, and fill out a two-page form. The court process



is slow, and a large corporation can easily drag out the proceedings while legal costs build up for an individual.

Winning a court case does not necessarily result in a windfall. The statutory damages are limited to \$100,000 for willful infringement (17 USC 504c). This could be bankrupting for an individual, but trivial to a corporation. Recovery for actual damages may be higher, but requires some proof of what those losses were, which can be particularly difficult in grey literature case, because of market value uncertainties. (17 USC 504b) The copyright owner can also recover any profits made by the infringer, but the court can take into account expenses and elements of profit not attributable to the infringement, which easily becomes an accounting nightmare.

### EXAMPLE ONE: IGNORING COPYRIGHT

In the late 1960s a political party known as the Black Panthers invaded campuses and television screens with their message of black power and independence. They wrote a wide variety of pamphlets to help spread their message. One of these was Eldridge Cleaver's "Credo for Rioters and Looters." (Cleaver, 1969). Another was Huey Newton's "Essays from the Minister of Defense." (Newton, 1967) These were small works, just a couple of pages long, typed, duplicated, definitely not printed on any sort of professional press. Party faithful handed them out on street corners, no charge. They represented a classic type of grey literature: self-published, free, intended to get across a message not make a profit. No one called them grey literature, though. Propaganda was the preferred term.

Copies of both of those pamphlets found their way to the Special Collections department of Michigan State University Library. This was somewhat unusual. They were nothing like Darwin first editions or the papers of a prize-winning author. It happened, however, that MSU Libraries houses the Russel B. Nye Popular Culture Collection, which includes American Radicalism. The pamphlets found their way into a vertical file, were cataloged, and became popular with students doing research for the freshman writing class called American Thought and Language. Most of the rest of the copies of those pamphlets went into waste baskets shortly after distribution. They had value to almost no one except the party faithful and scholars building a collection for the future.

In 1968, the operative copyright law in the US was still the old 1909 law. Copyright protection was not automatic, and even registration of the work meant nothing if all the copies did not contain a copyright notice of copyright. Leave off that notice, and any work which met the definition of being published fell immediately into the public domain, from which no rescue was possible.

No evidence exists for how the Black Panther Party, or Huey Newton, or Eldridge Cleaver actually analyzed the situation. But three suppositions seem reasonable. First, that the amount of trouble to register the pamphlets in accordance with the law seemed out of all proportion to the benefit. Second, that they actively wanted them copied, stolen, distributed as widely as possible, which copyright protection would only hinder. And third, they did not believe in the US government anyway. The latter was an outgrowth of their political views and not typical of grey literature publishers generally. The other reasons may have been more broadly shared: the Popular Culture Collection certainly has a large number of other pamphlets from the same and earlier eras without the required copyright notice -- all from organizations that explicitly just wanted their message to get out.

In a sense, the decision to eschew copyright did, in the end, benefit these two Black Panther pamphlets. Michigan State has been able to scan them and mount them on the Web in



digital form for students to use in research, and for anyone in the world with an internet connection to see. They have outlived the Black Panther Party, and their message is more widely available than ever.

If the Black Panther Party had published the same two pamphlets today, the legal situation would be very different. The lack of a notice and registration for the works would not matter. They would have automatic copyright protection. In fact there is no way that they could lose their protected status, whether they wanted to or not. Protection is both automatic and inescapable. If they wished people to have the freedom to reproduce and redistribute the pamphlets, they would have to take the proactive step of including an explicit permission in the work. In other words, it would require some minimal knowledge of US copyright law to achieve the same open access.

Although the Black Panther Party is gone, plenty of other groups are willing to pass out handbills, pamphlets, and other forms of grey literature similar in quality and quantity. This is particularly true around election time, and anyone who has seen the phenomenon has probably also seen the residue in nearby litter baskets. For this type of grey literature, the current copyright law has done nothing to increase its likelihood of retention. The message would go to the landfill, or to a paper-recycling plant, with perhaps one or two copies that a foresightful librarian files away in an acid-free folder in a vertical file waiting for the 95 year term for corporate, or the life-plus-70 term for authored works, to expire before access can legally be increased.

At present under US law, neither Newton nor Cleaver or their estates would have the moral right to withdraw the pamphlets because of changed views. Their choice to publish their views became immediately irretrievable.

#### EXAMPLE TWO: BELATED COPYRIGHT

On August 28, 1963 in Detroit, Michigan, Dr. Martin Luther King gave a speech which is now called the "I have a dream" speech. Not everyone classes speeches with grey literature, but they share many grey literature characteristics, especially when a text of the speech is distributed. King's speech is better known than the thousands of others that took place during those active years of the US civil rights movement, but it was not sold and not published by a standard commercial or university press. According to OCLC's WorldCat database, only about 15 libraries own the newsletter in which the text was distributed. Considering that this speech quickly became one of the most famous, and perhaps the most influential, in post-World War Two America, the scant listing is astonishing.

The speech and the newsletter copy ran into copyright problems almost immediately. King wanted a large audience for his speech. He encouraged the press to cover the event and to broadcast it nationally. Under the 1909 Copyright Act, King's speech counted as performance, not publication. King also had a history of registering his speeches for copyright protection, and in fact sent a copy of the speech to the Copyright Office for deposit on September 30, 1963. But the organizing committee had asked him to provide an advanced text or summary for the press. The copies distributed to the press had no copyright notice. Because of its impact, the speech was widely repeated and reproduced, including by a company called Mister Maestro, Inc. In a 1963 decision of the Southern District of New York, the court disagreed with the argument that the press copies had thrown the speech into public domain. It held that only a limited distribution of copies took place, not one which would constitute publication. The court recognized King's copyright, and assessed damages. (Nimmer, 1991, p. 159-167)



Nonetheless copies of the speech continued to be made and distributed, even though the King estate was relatively diligent about trying to enforce its rights. Few other owners of grey literature copyright have similar moral and monetary resources to fight infringements, though few grey literature examples are equally coveted. CBS was among those who included part (62%) of the speech in a recent documentary videotape, which they sold. Although the estate took them to court, a district judge reversed the prior ruling in 1998 and decided in favor of CBS, because he felt the circumstances surrounding the speech did in fact meet the definition of general publication. The ruling has been controversial and is currently under appeal. (Ringel, 1999)

However the appeal turns out, this example illustrates the kind of legal complications that can arise from the ambiguous publication status of a work of grey literature. Today the same set of problems could not be repeated under US law, because of the abolition of the notice requirement. But other similar ambiguities, such as unclear work-for-hire situations, exist. Ambiguities in general are more likely to involve grey literature because the urge to get information out, as with the King speech, often overrides any desire to make sure economic rights are protected.

In any case even the 1963 decision upholding King's copyright did not stop willful copying. People continued to act as if the speech were in the public domain.

### EXAMPLE THREE: EXPLICIT COPYRIGHT

Although desktop publishing software and laser printers made grey literature vastly easier to create in the 1980s, the growth of the Web in the 1990s turned almost every college student at a major university into a potential grey literature publisher, since most students have access to enough network-accessible storage to mount HTML files, and perhaps even a few scanned images. Much of this Web-based grey literature is the most ephemeral sort. It appears and disappears at the whim of its creator. Some of it may be for a particular class and vanish at the end of the semester. In any case universities tend to close the accounts within a year or so of graduation. Some students migrate their writings to a new home page on a commercial Internet Service Provider. But even when the contents remain unaltered, the address has changed.

One attractive example of this form of grey literature is Erica Olsen's home page "Es Chica Chica." [2] She set it up as an "ezine," and after several attempts, persuaded Yahoo to register it under "Entertainment > Humor > By Topic > Job Humor > Librarians" as "Why you Should Fall to your Knees and Worship a Librarian." Does registration with Yahoo make a site less grey? Perhaps, but the other problems of transience remain.

Unlike many Web publishers, Olsen has been careful to put copyright notices on her pages. She works in a research library, and has been accepted at library school, so she has a better than average understanding of copyright law. She knew that the law no longer required her to put the copyright notice on her Web page to give it protection. She added it only to warn visitors that she cared about protecting her material. (Olsen, 1999)

Her site has been popular, especially the three paragraph long "Why you Should Fall to your Knees and Worship a Librarian" section, and she maintains a guestbook for comments. In July, 1999, a visitor wrote:

Your website ROCKS!! We must have made 20 copies of "WHY YOU SHOULD...WORSHIP A LIBRARIAN" and passed them around to the 40+ librarians in our system. Gal, you're a welcome addition to our profession - come work in beautiful Long Beach, CA! -- McMillan, 1999



The visitor not only admits a violation, but seems to think that the infringement will please the copyright holder. Since the visitor appears to be a librarian herself, it may be reasonable to assume that she has a greater than average awareness of copyright law. She may well have behaved differently if Olsen's three paragraphs had appeared in a traditional print journal or in a regularly published book. But her attitude toward at least this form of grey literature seems to be that the copyright owner wants as much exposure for her message as possible. Her reasoning could easily be: access to the home page is free, so why should Olsen care about copies?

Olsen has told friends and colleagues about the infringement, but did not contact McMillan, although she had her email address. And she has not considered any legal action. This is typical of many grey literature infringements. The infringer probably really meant no harm, complaining would not undo the (largely moral) damage, and a lawsuit seems out of all proportion to her means or to the infringement itself. Instead, Olsen plans to be careful about putting any of her writings on the Web that she might later sell. (Olsen, 1999)

In practice, copyright offered no protection, even when handled with foresight and care.

### COPYRIGHT, GREY LITERATURE, AND THE NEXT MILLENNIUM

*Change in an ecology is systematic. When one element is changed, effects can be felt throughout the whole system. Local changes can disappear without a trace if they are incompatible with the rest of the system. -- Nardi, 1999, p. 51*

Will the casual attitudes toward grey literature, and infringing practices that accompany them, disappear in the new millennium? Two relatively recent events have changed the grey literature's ecological niche in potentially significant ways. One was the long-awaited US membership in the Berne Convention at the end of the 1980s. The other was the commercialization of the Internet in the mid-1990s. The former, in effect, extended protection to all grey literature produced in the US, and the latter created tens of thousands of virtual publishers who could market access to information without traditional publishers, bookstores, or any of the conventional bibliographic apparatus. At the least, Web publication has blurred the dividing line between grey and non-grey literature. Access to grey literature on the Web is currently not much different than, and often easier than, access to for-fee publications.

Web-based publishing has brought with it greater international access. This has exposed US grey literature to people in societies where the existence of both moral and economic rights is taken for granted. This exposure may ultimately increase respect for grey literature copyrights and reduce the readiness to infringe, especially if the philosophical basis for US copyright protection becomes less exclusively associated with economic value. But the process seems likely to be slow, in part because it conflicts with the urgent desire of scholars, librarians, and the creators of grey literature themselves to maximize accessibility. Infringement probably will not decrease until the creators decide to defend their rights. Until then, at least in the US, copyright for grey literature is almost meaningless. ]

### NOTES

[1] <http://www4.law.cornell.edu/uscode/17/>

[2] <http://www.msu.edu/user/olsencri/>



## REFERENCES

- Cleaver, E. (1969), *Credo for Rioters and Looters*, Black Panther Party. Available: <http://www.lib.msu.edu/spc/digital/radicalism/e185.61.e551969.htm>
- Foucault, M. (1973), *The Order of Things: an Archaeology of the Human Sciences*. NY: Vintage.
- McMillan, S. ( 31 July 1999), Message 226. Available: <http://www.guestbook.de/yasg.cgi?X=142268&I=2>
- Newton, H. (1968), *Essays from the Minister of Defense*, Black Panther Party. Available: <http://www.lib.msu.edu/spc/digital/radicalism/e185.615n481968.htm>
- Nardi, Bonnie A., and Vicki L. O'Day, *Information Ecologies: Using Technology with Heart*. MIT Press, Cambridge, MA, 1999.
- Nimmer, M., P. Marcus, D. Myers, D. Nimmer (1991), *cases and Materials on Copyright; and Other Aspects of Entertainment Litigation, including Unfair Competition, Defamation, Privacy*, 4th edition, West Publishing, St. Paul, MN.
- Olsen, E. (8 September 1999), conversation with Erica Olsen, East Lansing, MI.
- Ringel, J. (17 May 1999), "King vs CBS," *Connecticut Law Tribune*. Available on Lexis-Nexis.

# Intellectual property on the move

## Some observations on authors' rights, grey literature, publishers and last but not least access to information in a cyberian age

drs. Cees A.T. de Blaaij\*

### Introduction

The purpose of this article is to shed some light on the legal developments regarding intellectual property and grey literature in the digital environment.

First of all I want to discuss the current legal changes that are taking place in the copyright field and the electronic environment we call cyberspace, the electronic superhighway or any other overworked term which may refer to the conception of a computer networked world. Secondly I want to address the issue of authors' rights and grey literature in a networked information space.

I shall try to answer the question in what ways the current developments in copyright law will affect the electronic publishing and the access to grey literature. Also I would like to go into the role of commercial publishers and libraries in this matter.

### Pandora's Digital Box

This age we call the information society is the outcome of a transitional process from an industrial economy evolving into an information based economy. To put it more adequately: in this process of transition we make in general assumptions about the relevance of information often in the terms of the market and individual ownership. Copyright challenges not only our notions of property but our basic assumptions about the circulation of information in terms of production, distribution, accessibility and exploitation.

The introduction of digital technology and the coming of computernetworks linked to each other throughout the world created the fertile environment for fast and easy reproduction and almost unlimited distribution of an electronic work. These works include all kinds of copyright protected writings, images and sound recordings; all of them having one thing in common: the electronic format, patterns of ones and zeros.

This electronic format has the fluency properties of water and like a river heading to the sea electronic information flows from the creator to the enduser. However for some creators – e.g. those working in the music recording industry – the digital environment nowadays has become the modern version of Pandora's box because of its facilitation of unlimited reproduction and distribution of copyright protected digitized material all over the world.<sup>1</sup>

In this way the Internet has become a celestial jukebox for downloading music, software programs etc. for free without any consideration or – perhaps – knowledge about intellectual property rights. For copyright holders – like the entertainment industry, software producers etc. – the continuous infringement of their copyrights was a signal for starting a lobby campaign to protect their interests in the digital world. With some success. There was a lot of



pressure on legislative bodies for introducing and implementing new copyright laws tailored to the needs of copyright holders active in cyberspace.

### **Berne Convention and the TRIPs Agreement**

The Berne Convention is the foundation of contemporary international copyright law dating from 1886 and since then there have been many revisions<sup>2</sup>. Most countries did sign this convention. Some rather late like the US who signed the Convention in 1988. Most national copyright laws are based on the Berne Convention.

In 1994 the World Trade Organization (WTO) Agreement was signed together with an Annex to this called TRIPs – the Trade related aspects of Intellectual Property rights<sup>3</sup>.

Things gained momentum with the acceptance of the TRIPs Agreement which pushed the economic value of intellectual property by offering developed nations a more effective regime of protection in return for greater access to their markets on the part of commodity producers. But at the same time these producers also happened to be the consumers of products which were subject of intellectual property protection. The TRIPs Agreement strengthened in general the protection of intellectual property by supporting restrictions on freedom of access to knowledge and information.

Of course it is a strange combination when on the one hand the freedom of trade is stimulated by breaking down trade barriers and on the other hand restrictions on the access to knowledge and information are enforced.

### **The WIPO Copyright Treaty**

With the World Intellectual Property Organization Conference in December 1996 there were a number of far reaching proposals on the extension and expansion of copyright which were finally rejected in the adopted version of The WIPO Copyright Treaty<sup>4</sup>.

The WIPO Copyright Treaty is a special agreement under Article 20 of the Berne Convention<sup>5</sup> and grants to authors more extensive rights than the Convention. The rejection of the original draft was of great importance because the draft was heavily weighted against public and user interests.

The proposals regarding to the right of reproduction and the right of communication in Articles 7<sup>6</sup> and 10<sup>7</sup> of the original draft WIPO Copyright Treaty were moved forward just a few months before the Conference started. This narrow time schedule reduced the opportunity for national countries to give comment and consult with user groups for reaching a kind of consensus.

### **Right of Reproduction**

If article 7 had been accepted then this would mean that all electronic copies no matter how temporary were reproductions. In practice this would make copies made in a computer memory or shown by means of a browser on a screen an infringement unless permitted by law or by the copyright holder.

Instead of article 7 an agreed statement was accepted which stated that the reproduction right as formulated in Article 7 did fully apply in the digital environment. But at the WIPO conference no binding decision was made regarding the reproduction right which covers temporary copies.



Storage on a hard disk or in a computer memory is a particular mode of temporary or transient fixation. But the definition also depends on the type of technology used. And as we all know technology is dynamic by nature. For example ways of storing information by means of caching may become outdated because there will be other and more sophisticated solutions for storing electronic information. This means also that original article 7 in this case would become obsolete.

This shows even more that we don't know enough from the technological problems we may encounter in the future. Therefore, from a technological point of view it would be premature to implement the original article 7.

The adopted WIPO Copyright Treaty grants to authors more extensive rights. An extension of the period for protection of works was adopted to life plus 70 years. In my opinion no real arguments can be found for justifying this adoption. It is really no incentive for authors to produce works that otherwise will not be produced. The extension benefitted the publishers because the extension resurrects copyright in some works after the old period of protection has expired.

The adoption of the treaty was heavily influenced by an intensive lobbying campaign from the side of large commercial publishers and information providers. They pushed for a copyright reform centered on the extension of copyright protection and the expansion of the concept of copyright. This position is being dictated by the fear of what unforeseen consequences may follow in a world of fast changing technology.

This is also visible in the U.S. Report on the National Information Infrastructure<sup>8</sup> and the E.C. Green Paper on Copyright and Related Rights in the Information Society<sup>9</sup>. Also the European Commission made a recent proposal for a Directive<sup>10</sup> which results in the strengthening of the position of copyright holders interests.

Libraries and researchers oppose that transient copying is infringement because they want to apply the same rights of users which are valid in the analog world. Publishers on the other hand fear that approving temporarily stored electronic information will open Pandora's digital box. From their point of view it will be better to protect any type of temporarily stored information to prevent any erosion at all. This is what I call the Fort Knox<sup>11</sup> approach. But the question is still open and has to be resolved within the context of Article 9<sup>12</sup> of the Berne Convention if temporary copying constitutes an infringement.

### **The Digital Millenium Copyright Act**

The implementation of the WIPO Copyright Treaty in the US has resulted into the introduction in October 1998 of The Digital Millenium Copyright Act<sup>13</sup> (DMCA) to meet the conditions which were formulated in the WIPO Copyright Treaty.

This new addition to American copyright law is very complex by nature. The DMCA gives the greatest protection to copyright owners' right to control access to information since it make it a violation both for users to circumvent access controls, and for others to manufacture, distribute or offer devices that circumvent technological measures as means of protection of copyrights.

The DMCA leaves some uncertainty for the public interest because Congress delayed implementation of the prohibition on circumvention of access controls for a two-year period<sup>14</sup> and further instructed the Librarian of Congress in consultation with the Register of Copyrights, to identify particular classes of works whose users who would be "adversely



affected by the prohibition...in their ability to make noninfringing uses under this title of a particular class of copyrighted works<sup>15</sup>. It is not clear if these classes of works include scientific journal articles and other materials like grey literature which are relevant to non-profit purposes.

In general you get the impression that the DMCA in its final appearance is systematically incoherent. All those involved in the negotiations got a piece of the copyright pie but in a way that resulted in securing certain extremely specific exceptions, rather than mentioning general limiting principles. They did not look or rather failed to look at the broad picture. As a result it is uncertain how the balance between owners and users will evolve in the future. Taking the developments in European Union into account the overall picture doesn't seem very bright for users and libraries.

### **The Nature of the Copyright Balance**

Another important question is raised in what ways constitutional rights are harmed with these new forms of digital copyright protection. There may be a point that it may be necessary to bring copyright in balance with the U.S. First Amendment and Article 10 of the European Convention on Human Rights.

I refer here to the U.S. Constitution. In Article 1, § 8, clause 8 confers upon Congress the power to "promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." The copyright protection is derived from this clause. The author is granted exclusive rights by means of the Copyright Act to promote useful arts and sciences by securing for the life-time of the author plus 50 years of several exclusive rights with regard to their writings.

So the exclusive rights only exist to the extent they are necessary as an economic incentive for creation. There is no legal or moral basis in the American Constitution for authors to have complete control over the works they create<sup>16</sup>.

If we look at copyright history then we see that this thought already originated within The Statute of Anne, passed in England in 1710 and a landmark in the history of copyright law. It recognized that authors should be the primary beneficiaries of copyright law; it also established the idea that such laws should have only limited duration, after which works could pass into public domain. The designated period came to be set at 28 years.

So what it is stated within the US Constitution is not new. Congress made it ninety years already clear that the interest in protecting the rights of authors is secondary to the interest in maintaining public access to creative works<sup>17</sup>. Also article 10 from the European Convention on Human rights guarantees this type of freedom of communication and information.

The exclusive rights from the copyright holder exists as a tool and are necessary as an economic incentive for creation. So the conclusion is that there must be rights sufficient to provide an incentive to create new works but not to an extent that hinders access to works. Both in the European Union and in the US there is a constitutional basis for this conclusion.

### **Keeping the Balance**

Without paying enough attention to the rights of public access and use and at the same time not reaching a general consensus in society, the new developed copyright legislation therefore serves the already established few who can get them passed and does not benefit society as a



whole.

In the preamble of the WIPO Copyright Treaty there was a Recital which recognized the need to keep the balance between the rights of authors and 'the *larger* public interest', particularly education, research and access to information. In the original draft Treaty there was no such Recital. So it is important that the Recital exists and confirms the primacy of the public interest as a goal to the dissemination of knowledge and free access to information and ideas. It is necessary that a better tuned legal framework has to be developed that can be used to balance the proprietary and commercial interests with all parties involved.

As already stated before, the larger public interest is constitutionally protected in international jurisdictions by a guarantee of freedom of communication including the right to disseminate ideas and information and have access to them.

Intellectual property law contains numerous balances between the right of creators against the rights of the public at large in a rich public domain but overprotection of copyright holders – the Fort Knox approach – won't benefit either user nor creator in the digital world. In the words of judge Kozinski:

"Overprotecting intellectual property is as harmful as underprotecting it. Creativity is impossible without a rich public domain. Nothing today, like nothing we since we tamed fire, is genuinely new: culture, like science and technology, grows by accretion, each new creator building on the works of those who came before. Overprotection stifles the very creative forces it supposed to nurture"<sup>18</sup>

The right of access to information in the electronic environment is just as fundamental as protecting the intellectual property rights of copyright holders.

### A new direction ?

The digital environment makes it necessary to adapt existing copyright law to new technological circumstances but one can ask if traditional intellectual property ideas and values from the offline, physical world will also work in the legal field online world.

According to John Perry Barlow<sup>19</sup> - one of the founders of the Electronic Frontier Foundation<sup>20</sup> - we have to do with a profoundly new kind of challenge.

In his view the copyright law from the physical world cannot be applied to a digital environment because in his words "the accumulated canon of copyright and patent law, was developed to convey forms and methods of expression entirely different from the vaporous cargo it is now asked to carry"<sup>21</sup>.

Should we as Barlow observes it 'dance on the grave of copyright'<sup>22</sup>. This will solve very little according to Barlow 'especially when so few are willing to admit that the occupant of this grave is even deceased and are trying up by force what no longer can be upheld by popular consent'<sup>23</sup>.

In my opinion Barlow's prophecy is a too premature in announcing the death of intellectual property. History proves that old media are never fully swallowed up by newer media so intellectual property will be there to stay, even in the digital world. Notwithstanding Barlow's judgment on the future of copyright his arguments show us that we need a different approach how to deal with Pandora's digital box.

Compared to the off-line world the networked environment creates totally new possibilities for communication and information distribution all over the world. Information has become a

commodity in an economic world where breaking down trade barriers and the promotion of free world trade is the underlying driving force.

As a kind of case study for new models of exploitation rights I will examine the situation of grey literature in an networked space and intellectual property issues related to this.

### **The case of scientific e-publishing**

There are two types of grey literature which have an important share in the area of grey publishing in general. On the first place scholarly information and secondly government information<sup>24</sup> in different kinds of format: working papers, policy statements, reports etc. A fairly new development is the substantial amount of grey publishing currently carried out on the Internet by a number of scholars and researchers without the involvement of commercial publishers.<sup>25</sup> The way Internet offers new possibilities made many people realize that the commercial model is not very comfortable to the need to have free circulation of ideas among scientists. And that is what is essentially about: scientific communication is a matter of sharing ideas and have free access to a common pool resource.

Typically enough this is the second time this happens in history because the birth of the scientific journal itself in the XVIIth century was also an answer to the needs of scientists to have a communication platform to stay in touch with each other.

*Mutatis mutandis* the electronic journal and e-print archives also pick up the basic idea of sharing information and scientific knowledge. And this comes very close to the objective of copyright as formulated in US Constitution, the European Convention of Human Rights and more specifically in the Recital of the WIPO Copyright Treaty pointing at the digital environment.

### **Commercial scientific publishing**

Commercial publishers entered the field of the scientific journal from the 19th century onwards. After the Second World War, commercial publishing became the dominant mode of distribution. Nowadays only a small number of international publishers like Reed-Elsevier, Wolters-Kluwer, Springer and several others have established themselves as the most important players in the business of distributing scientific journals.

#### ***Profits and prices***

Commercial publishing has become big business. The mergers and acquisitions which have taken place and continue to occur with great frequency are but a reflection of what's happening in the business and industry. The large publishers go for market share, industry penetration and profits. It is no secret that Reed-Elsevier enjoyed high returns on its stock. Reed-Elsevier reported profits of \$ 378 million on sales of \$ 938 million in its scientific activities alone. It is also no secret that Reed-Elsevier sold its consumer magazines publishing division so it could focus on scientific/technical journals with a higher profit-margin. Reed-Elsevier also wants to be the dominant publisher in scientific/technical journals market. Some fifteen hundred titles are listed in the current Elsevier Science: 1999 Subscription Price List. Prices range from a modest \$ 6500 for an annual subscription to *Analytica Chimica* to a nice price of *Excerpta Medica* for \$ 49000. Subscriptions to all journals would cost between 1.6 million and three million a year and it still goes up.



*Commercial publishing competitive and value-adding ?*

Another question is if the commercial publishing industry is competitive enough as a mode to produce at the least costly mode. If we look at the amount of profits of some commercial publishers then this remains to be seen. Many feel that publishers of high-price, small circulation journals are making excess profits. Can somebody justify the doubling of the cost of Elsevier's Brain Research between 1992 and 1996 to \$ 15,000 annually ?

What can fairly be expected ? On basis of earlier arguments I would welcome an investigation to anti-competitive practices in the scientific publishing sector from the US Government or in the European Union. The reason for this is on the first place to get a balanced view on the whole matter of scientific publishing and secondly to correct the threatening monopolization of scientific information by stimulating competitive practices in scientific publishing. So in the end we can pay a reasonable price for information in an area where information needs to be shared between users according to the formulated constitutional principles of scientific information.

Do high subscription rates also justify the quality of commercially published scientific journals ?

The production process of a journal consists of authoring, editing, presentation, duplication, distribution, promotion, integrity assurance and branding. Each of these activities may contribute to a higher value-added product and are essential features of paper publications. Are they worth the cost ? The cost consist of the amount of money involved in producing the journal and the delay in publication creating a loss of currency. So the only way to judge the value would be to enhance the competition between modes of publishing to reach an affordable price level.

*Serial crises*

Why do they charge so much for what is essentially a printing and distribution service ? The answer is simple. They can do it because the market allows it. Prices rise to the maximum level the customer can bear with catastrophic results for academic libraries in general. The US Association of Research Libraries calculated that 114 member libraries spent 142 per cent more on journals in 1997 than ten years before, but ordered 6 per cent fewer titles. This is what is called the 'serial crises'.

*Role of libraries*

Many academic libraries, at least in my country and many other countries in Europe, could not cope with the phenomenon of steep rising prices and canceled subscriptions to a lot of scientific journals. This way of economic exploitation of scientific information has lead in some cases toward a situation of information poverty, in which rights of accessibility for students and teachers are violated in the perception of scientific information as a public good. In this way a fundamental relation exists between the freedom of information and the exploitation of intellectual property rights.

Publishers like the traditional journals because take care for a steady income flow. For libraries it is a totally different world. They are left to the mercy of those who eat their budgets, so it is quite natural from the perspective of survival that libraries cancel subscriptions to particular journals, while on the other hand making sure that some other library in the same area continues to subscribe. In this way sharing arrangements are made. The cancellation of subscriptions reduces the income for the publishers and 'forces' them – they say – to increase the subscription prize. This produces the countereffect: libraries will



continue to cancel subscriptions. So is there any hope left for libraries ?

### *The library 'empire' strikes back*

In my opinion the end is approaching in what libraries can endure. Libraries in the US and Europe are going to strike back, the ultimate goal being to return control of scholarly publishing to the non-profit societies and what they consider to be responsible publishers. Then again this supports the idea of a more free circulation of scientific information.

### Can the circle be broken ?

#### *Feasibility scientific publishing: the pioneering role of grey literature*

Is scientific publishing on the Internet without the involvement of commercial publishers feasible?

As stated before more competition between modes of scientific publishing is welcome as a matter of lowering costs and at same time reach an acceptable level of quality.

The way how grey literature is being produced and published on the Internet could serve as a prototype how things might work successfully for electronic scientific publishing in general. The publishing of grey literature in the digital environment sets an example for other categories of electronic publications.

In my opinion even in the paper society grey literature was more successful in communicating novel ideas mainly through direct communications from person-to-person: the use of the human networkmodel. Also the use of working papers or conference papers contributed to a relative efficient circulation of scientific information at low cost.

Electronic publications also offer the opportunity of experimentation to get away from a number of limitations which are intrinsic to paper publishing.

There are number of advantages. The most important is currency advantage. Then there is the relatively low cost of producing an electronic journal in comparison to a traditional journal. It has no printing cost and no distribution cost per copy put in circulation.

An important feature is the quality of 'gate-keeping' a security control to ensure the quality and going building on the status 'branding' of a journal. This is a difficult issue for electronic journal to tackle and also in relation to the aspect of asserting integrity assurance. How are electronic materials going to be archived for posterity ?

The peer reviewed status of important paper based scientific journals is still a threshold. Many researchers with knowledge about E-journals still favor in the end the traditional journals with a good reputation. Because this is what still counts for researchers: the number of publications in the established journals.

#### *Position researchers*

Also the position of researchers as a group cooperating with commercial publishers is at stake. Researchers depended in the pre-Internet time heavily on the services of publishers and in return for that they demanded exclusive copyright from authors. This allows the publisher to control and charge for any subsequent use of the published material – even by the authors themselves. Scientists also depend on publishing for career advancement so they don't have an incentive to stop submitting articles. Afterwards the material can be reprinted or even electronically reused. This is a very nice deal for the publisher who pays none of the costs of originating his material and gets paid by the same libraries of institutions that are feeding the publishers.



*E-print vs. P-print: a undecided case*

E-print archives will be a model to follow from scientific researchers and responsible publishers. However a number of conditions will have to be met to be successful like I mentioned before.

The contents of an electronic publication needs to be careful organized to help the reader find the relevant information. Also the peer review of submissions is something which can't be missed. Last but not least the integrity needs to be assured.

It is interesting to know if e-journals have become accepted as a new tool for distributing scientific information. Despite library support the success of electronic publishing is not guaranteed. Electronic journals face the same difficulties to those of any new journal in getting established and attracting authors. It's also a matter of competing with similar printed publications. That doesn't mean that both types of journals can't be living together. In physics, the Los Alamos e-print repository – where preprints are submitted by scientists and are available for free – millions of transactions are made. These preprints have supplanted journals in distributing primary literature but didn't have a negative financial impact on the printed physic journals.

*Two models of scientific publishing and grey literature*

The picture which emerges is that there are two kinds of models of scientific publishing. The first model is the way in which grey literature is electronically produced. This is an idea-sharing model that benefits researchers in their exchange of knowledge and information and in public interest. Information is seen as a free good to benefit the quality of scientific publishing.<sup>26</sup> At the same time the idea-sharing model supports the notions about the scope and the objective of copyright as formulated within the US Constitution and the Recital of the WIPO Copyright Treaty

The commercial model could play a role in scientific publishing insofar a successful title from the online world would be transferred to a commercial publisher. A drawback would be the immediate substantial price increase. So the commercial model does the same in the online world as in the offline: just pay for what you read.

The question is what the future will be for grey literature in relation to publishing and commercial publishers. Do commercial publishers want to play an intermediary role also in the grey literature field in some way or another without turning the Internet into an electronic Fort Knox by means of taking over from authors their intellectual property rights ?

Their position seems difficult owing to themselves for digging in and defending their traditional publishing stronghold by lobbying for strengthen intellectual property rights. What they do is applying a kind of military strategy by not only protecting contents but also trying to control access to information. This is a new development but in my opinion from a strategically point of view a strange choice. This sound almost like the swan song, the farewell speech for traditional publishing. I would rather suggest if you can't beat the idea sharing people, join them !

It could very well be that the large commercial publishers are at end of their lifecycle and just don't know how to respond to the new challenges of the digital world.

The production of grey literature has set such an example and shows a method for a viable

way of publishing scientific information. One concern is still luminent. Large companies with more money could squeeze not-for-profit publishers out of the electronic journals market. This can only be solved by counterbalanced measures which would stimulate competitive practices. Perhaps the academic world in the US or Europe could vote for an agency ensuring principles of a competitive market but also protecting the interests of the scientific community in sharing information.

At this moment there are a number of not-for-profit setups to help universities and societies to publish at low cost like High Wire, an initiative by Stanford University Libraries and Academic Information Resources. It will take probably a number of years to reach a new equilibrium in the market before the not-for-profit approach is rewarding enough

The Scholarly Publishing and Academic Resources Coalition (SPARC), set up in 1997 by the US Association of Research Libraries, is even more determined to engage the battle with commercial publishers by launching journals which are aimed at competing with equivalent expensive titles. Libraries who have teamed up with SPARC promised to buy each title<sup>27</sup>

The Internet promises – or threatens perhaps – to change fundamentally what we mean by a book or a journal. And also the relationship among author, publisher and reader will change. Instead of a short term approach for implementing solutions to the opening of what looks like Pandora's digital box a more long term approach seems to be advisable to solve intellectual property issues.

As we had not much practical experience with the electronic media in general, it takes time to find a useful strategy how to operate legal intellectual property instruments in an electronic environment.

The objectives for a long term strategy are in my opinion twofold. Firstly we ought to adapt copyright law itself to new technological circumstances. Not by imposing old rules which are valid in an offline world and which are based on physical expression, but we should respond with new online rules in protecting copyright holders' interest in their creation which do justice to electronic works.

Secondly at the same time we should allow users to benefit from new technology. This adaptation is crucial to maintain the balance between the rights of authors and the public. Copyright law itself doesn't just apply to protection of authors but it has also a social value.

The working of the present intellectual property laws are based on territorial principles. So each country has its own set of 'rules'. So it is important to reach an international consensus on the scope of copyright. Intellectual property rights could be considered as kind of property rights reserved to the owner. Commercial publications rely on the private property mode, however not in a way this would obstruct the sharing of information as ideas. So the access to information would have to be better guaranteed for certain groups of users: students and researchers and teachers.

Uses of grey literature fall for a part in the public domain and might as well be given free away. Other parts of scientific publishing on the Internet deserve others forms of protection which conforms to the common property model. Common property model doesn't mean the absence of property. Authors of scientific works on the Internet can allow free use of their work for scientific communication by others, with reserving the right for republishing for commercial reasons.

This leads to the conclusion that the mentioned models of publishing can live together each



with his own set of rights. Only time will tell us which will survive in the end.

To end my paper I will tell you something about the situation concerning scientific publishing in the Netherlands. There the discussion goes on between scientists and universities regarding the possession of intellectual properties and the advent of the Internet.

The Universities claim that they are better equipped as an organization to make collective bargains and there therefore will be able to protect the interest of individual scientists. So universities expect that scientists will transfer their rights without any reward for this. Law by nature doesn't work in this way. Of course universities offer some advantages eg. scale, setting up consortia, negotiating licence agreement with publishers etc. But just giving away the rights by a scientist would only reduce the incentives for the scientist and harm the so called academic freedom. A better alternative would be to negotiate on basis of contract law, with properly rewards for the author being build in the contract.

The best option however would be if the scientists decide for themselves which mode of communication they wanted to follow and taking the influential role of grey literature into account – which blossoms outside the commercial sphere - the common property model wouldn't be a bad choice. Under common property the publication can be used by a particular group of persons, but is closed bu outsiders.

As scientific information is considered as kind of limited public good there have to be restrictions about how and how much each person may use it. Within these restrictions the information should be open to all members.

Thank you for willing and kind attention.

Autobiographical note:

drs. Cees de Blaaij is law librarian and head of the Documentation Center from the Institute for Information Law, University of Amsterdam, Netherlands.

Address:

University of Amsterdam  
Institute for Information Law  
Rokin 84  
1012 KX Amsterdam  
Netherlands  
tel. 31-20-5253927  
fax 31-20-5253033  
email: blaaij@jur.uva.nl or cees@deblaaij.net

<sup>1</sup> The copying and abundant distribution of music recordings in a digitized, compressed format; MP3 throughout the Internet is one of the striking examples of the impact of a digital technology

<sup>2</sup> Noteworthy is the widening of the scope of copyright protection through the years since the Berne Convention came into existence.

<sup>3</sup> Available at <<http://www.wto.org/wto/intellect/1-ipcon.htm>>

<sup>4</sup> WIPO Diplomatic Conference on Certain Copyright and Neighbouring Rights Questions, Agreed Statements Concerning the WIPO Copyright Treaty (Dec. 23, 1996) <<http://www.wipo.org/cng/diplconf>>

<sup>5</sup> Berne Convention for the Protection of Literary and Artistic Works, Paris Act 1971. WIPO, Geneva, 1989

<sup>6</sup> Article 7: Scope of Right of Reproduction. Article 7(1): The exclusive right accorded to authors of literary and artistic works in Article 9(1) of the Berne Convention of authorising the reproduction of their works shall include direct and indirect reproduction of their works, whether permanent or temporary in any manner or form.

<sup>7</sup> Article 10: Right of Communication. Without prejudice to the rights provided for in Articles 11(1)(ii), 11bis(1)(i), 11ter(1)(ii), 14(1)(i) and 14bis(1) of the Berne Convention, authors of literary and artistic works shall enjoy the exclusive right of authorizing any communication to the public of their works, including the making available to the public of their works, by wire or wireless means, in such a way that members of the public may access these works from a place and at a time individually chosen by them.

<sup>8</sup> Information Infrastructure Task Force, Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights 7 (1995) [also White Paper]

<sup>9</sup> EC Green paper 'Copyright and Related Rights in the Information Society', COM (95) 382 final, of 19 July 1995

<sup>10</sup> Proposal for a Directive on Harmonization of certain aspects of Copyright and Related Rights in the Information Society, COM (97) 628 final

<sup>11</sup> For maximum security the U.S. Bullion Depository, a solid square bomb-proof structure with mechanical protective devices, was built there in 1936 to hold the bulk of the US nation's gold.

<sup>12</sup> Berne Convention: Article 9 [Right of Reproduction: 1. Generally; 2. Possible exceptions; 3. Sound and visual recordings] (1) Authors of literary and artistic works protected by this Convention shall have the exclusive right of authorizing the reproduction of these works, in any manner or form. (2) It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author. (3) Any sound or visual recording shall be considered as a reproduction for the purposes of this Convention.

<sup>13</sup> Pub. L. No. 105-304, 112 Stat. 2860 (1998) (adding §§ 512 and 1201-05 to the Copyright Act of 1976)

<sup>14</sup> See 17 U.S.C. § 1201(a)(1)(A)

<sup>15</sup> *Id.* § 1201(a)(1)(C).

<sup>16</sup> *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 428 (1984)

<sup>17</sup> "The enactment of copyright legislation by Congress under the terms of the Constitution is not based upon any natural right that the author has in his writings, ... but upon the ground that the welfare of the public will be served and progress of science and useful arts will be promoted by securing to authors for limited periods the exclusive rights to their writings" (H.R. Rep. No. 60-22222, at 7 (1909)).

<sup>18</sup> *Vanna White vs Samsung Electronics America Inc*, 989 F.2d 1512, 1513, 26 USPQ 1362 (Kozinski, J. dissenting).

<sup>19</sup> John Perry Barlow is a retired cattle rancher, lyricist for the Grateful Dead, co-founder of the Electronic Frontier Foundation and professor at The Berkman Center for Internet & Society at Harvard Law School



<sup>20</sup> EFF, the Electronic Frontier Foundation, is a non-profit organization working in the public interest to protect civil liberties, including privacy and freedom of expression, in the arena of computers and the Internet. EFF was founded in 1990.

Their website is <<http://www.eff.org>>

<sup>21</sup> Barlow, J.P., 'Selling Wine without Bottles. The Economy of Mind on the Global Net, in: P. Bernt Hugenholtz (ed.), *The Future of Copyright in a Digital Environment*, Den Haag : Kluwer Law International, p. 169,170.

<sup>22</sup> Id. p 174

<sup>23</sup> Id. p.174

<sup>24</sup> As a democracy becomes more online and citizens are stimulated by governments to participate in the information society they will get more involved in the political process. From that perspective it is considered essential that government information is being put into the public domain and people will have access to relevant information.

National parliaments stimulate or have been stimulating this process by introducing supportive legislation like the Freedom of Information Act in the US and the policy guidelines from the European Commission for the establishment of the Information Society project.

In this way a part of grey literature produced from all kinds government agencies is in the public domain free from copyrights and very often electronically available on the Internet.

<sup>25</sup> Farace, Dominique J., *Rise of the Phoenix: A Review of New Forms and Exploitations of Grey Literature*, in: *Publishing Research Quarterly*, vol. 13, nr. 2 (1997)

<sup>26</sup> John Perry Barlow defended this model in his famous article J.P. Barlow, 'The Economy of Ideas. A framework for rethinking patents and copyrights in the Digital Age, in: *Wired*, 2.03, 1995

<sup>27</sup> *Nature* 393, 719; 1998