

Nineteenth International Conference on Grey Literature

Public Awareness and Access to Grey Literature

National Research Council of Italy, CNR Rome • October 23-24, 2017



Program Book

ISSN 1385-2308

Conference Host and Sponsors:



Inist



SPRINGER NATURE



WILEY



GL19 Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
www.textrelease.com • conference@textrelease.com
Tel/Fax +31-20-331.2420



CIP

GL19 Program Book

Nineteenth International Conference on Grey Literature – Public Awareness and Access to Grey Literature - Rome, Italy, on October 23-24, 2017 / compiled by D. Farace and J. Frantzen ; GreyNet International, Grey Literature Network Service. – Amsterdam : TextRelease, October 2017. – 124 p. – Author Index. – (GL Conference Series, ISSN 1386-2308 ; No. 19).

TIB (DE), DANS-KNAW (NL), FEDLINK-Library of Congress (USA), CVTISR (SK), EBSCO (USA), Inist CNRS (FR), ISTI-CNR (IT), KISTI (KR), NIS IAEA (AT), NTK (CZ), and The New York Academy of Medicine (USA) are Corporate Authors and Associate Members of GreyNet International. This program book contains the schedule for the plenary and panel sessions, as well as the poster session and sponsor showcase. The titles and abstracts of the papers as well as information on the authors are provided. When available, copies of the PowerPoint slides are included in notepad format.

Foreword

PUBLIC AWARENESS AND ACCESS TO GREY LITERATURE

Two of the most formidable problems that have faced information through the years are its overload on the one hand and its loss on the other. These are seen as interconnected with the supply and demand sides of grey literature.

A quarter century ago, the Grey Literature Network Service joined by research communities in library and information, physics, karst and marine sciences, biomedicine, nuclear energy, archeology, and many other scientific and technical fields set out to address this loss and overload of information.

In 1992, when the call for papers went out for the first conference in the GL-Series, the response was predominantly focused on the demand side of grey literature – that which was difficult to find and even more to access. The emphasis then lie in stemming the loss of grey literature. However, the outcome of that first conference also called attention to the equally important need for further research into the supply side of grey literature – namely its production, publication, and public awareness.

GL19 seeks to demonstrate how researchers and authors in the last 25 years have made significant inroads in responding to the loss and overload of grey literature. Likewise, this conference will seek to provide new directions in achieving public awareness and access to grey literature on an ever changing information landscape.

Dominic Farace
GREYNET INTERNATIONAL

Amsterdam,
OCTOBER 2017

GL19 Conference Sponsors



Biblioteca Centrale 'G. Marconi', Italy
National Research Council of Italy, CNR



ISTI, Italy
Institute of Information Science and Technologies
National Research Council of Italy, CNR



The New York Academy of Medicine, USA



KISTI, Korea
Korea Institute of Science and Technology
Information



CVTISR, Slovak Republic
Slovak Centre of Scientific and Technical Information



EBSCO, USA



NIS-IAEA, Austria
Nuclear Information Section;
International Atomic Energy Agency



TIB, Germany
German National Library of Science and Technology –
Leibniz Information Centre for Science and
Technology University Library

GL19 Conference Sponsors



DANS, Netherlands
Data Archiving and Networked Services;
Royal Netherlands Academy of Arts and Sciences



NTK, Czech Republic
National Library of Technology



FEDLINK, USA
Federal Library Information Network;
Library of Congress



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



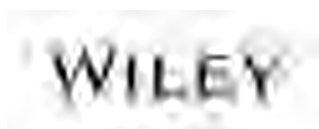
ACM Digital Library, USA
Association for Computing Machinery, Inc.



Springer-Verlag GmbH, Germany



EIPASS, Italy
European Informatics Passport



John Wiley & Sons, Inc., USA

Special Thanks

Conference Planning and Logistics • Rita Ciampichetti and Luisa De Biagi

GL19 Program Committee



Alberto De Rosa ^{Chair}
National Research Council of Italy
CNR Central Library 'G. Marconi'
Italy



Stefania Biagioni
Institute of Information
Science and Technologies
ISTI-CNR, Italy



Margret Plank
German National Library
of Science and
Technology, Germany



Ján Turňa
Slovak Centre of Scientific
and Technical Information
CVTISR, Slovak Republic



Meg Tulloch
FEDLINK
Library of Congress
United States



Petra Pejšová
Wikimedia WM CZ
Czech Republic



Dobrica Savić
Nuclear Information
Section, International
Atomic Energy Agency,
Austria



Christiane Stock
Institut de l'Information
Scientifique et Technique
CNRS, France



Hana Vyčítalová
National Library of
Technology,
Czech Republic



YongJu Shin
Korea Institute of Science
and Technology
Information, KISTI
Korea



Deborah Rabina
Pratt Institute,
School of Information
United States



Tomas A. Lipinski
University of Wisconsin,
Milwaukee, UWM
United States



Joachim Schöpfel
University of Lille
France



Dominic Farace
GreyNet International
Grey Literature Network
Service, Netherlands

TextRelease

GL19 Program and Conference Bureau

Table of Contents

	Foreword.....	3
	Conference Sponsors.....	4
	Program Committee	6
	Program Chair and Conference Moderators	8
	Program Outline	9
	Conference Program and Timetable	10
<i>Program</i>	Opening Session	14
	Session One – Exposing Grey Literature to Wider Audiences.....	23
	Session Two – Overcoming Obstacles in Accessing Grey Literature.....	37
	Poster Session and Sponsor Showcase.....	60
	Panel Session – Innovations in Grey Literature Powered by Research Data.....	79
	Session Three – Impact of Emerging Technologies & Social Media on Grey Literature.....	103
<i>Advertisements</i>	ISTI-CNR, Institute of Information Science and Technologies.....	16
	FEDLINK, The Federal Library and Information Network - Library of Congress.....	36
	The New York Academy of Medicine.....	48
	CVTISR, Slovak Centre of Scientific and Technical Information.....	54
	NTK, National Library of Technology, Czech Republic.....	68
	KISTI, Korea Institute of Science and Technology Information.....	74
	INIS, The International Nuclear Information System	76
	LISTA-Full Text EBSCO.....	78
	DANS Your 7 steps to sustainable data.....	78
	PsycEXTRA EBSCO.....	102
	GL20 Pre-Conference Announcement.....	121
	GL19 Conference Proceedings Order Form.....	122
	GL19 Conference Proceedings Order Form.....	122
<i>Appendices</i>	List of Participating Organizations	120
	Index to Authors.....	123



Moderator Day One

Margret Plank

Head of Development

German National Library of S&T

Margret Plank is currently the Head of Development and the Competence Centre for Non-Textual Materials at the German National Library of Science and Technology in Hannover, Germany. The aim of the Competence Centre is to develop emerging tools and services that actively support users in the scientific work process enabling non-textual material such as audiovisual media, 3D objects and research data to be published, found and made available on a permanent basis as easily as textual documents. Previously she was responsible for Information Competence and Usability at the TIB. She has also worked as a researcher at the Institute of Information Studies and Language Technology of the University of Hildesheim. She represents TIB on a number of boards including IFLA Steering Committee Audiovisual and Multimedia Section as well as the International Council for Scientific and Technical Information, ICSTI/ ITOC.

Email: margret.plank@tib.eu



Program Chair

Alberto De Rosa

Head Central Library

National Research Council of Italy

Alberto De Rosa is responsible for the National Research Council Central Library 'G. Marconi', the main Italian multidisciplinary library devoted to Science and Technics. He is involved in technological, administrative, and management activities related to scientific projects and Information Science in both National and European programs. He graduated in Economics at the Naples University 'Federico II'. He is a chartered accountant and statutory auditor (Register of the Italian Ministry of Economics and Finance). He obtained various Masters in Business and Management in the Public Administration and Research boards. From 1993 to 2013 he has been Administration Manager of CNR Research Institutes (e.g. the Institute of Biostructures and Bio-imaging); From 2002 to 2013 he has been Adjunct Professor of 'Company Structure' at the Naples University. And, from 2006 to 2014 he was Responsible for document management system, document flow and current archives at the CNR Institute of Biostructures and Bio-imaging.

Email: alberto.derosa@cnr.it



Moderator Day Two

Judith C. Russell

Dean of University Libraries

University of Florida

Judith C. Russell is the Dean of University Libraries at the University of Florida. She was formerly the Managing Director, Information Dissemination and Superintendent of Documents, at the U.S. Government Printing Office (GPO). Russell previously served as Deputy Director of the National Commission on Libraries and Information Science (NCLIS) and as director of the Office of Electronic Information Dissemination Services and Federal Depository Library Program at GPO. She worked for more than 10 years in the information industry in marketing and product development, as well as serving as a government-industry liaison. Her corporate experience includes Information Handling Services (IHS) and its parent company, the Information Technology Group; Disclosure Information Group; Lexis-Nexis (former Mead Data Central), and IDD Digital Alliances, a subsidiary of Investment Dealers Digest. She has an M.L.S. from Catholic University and a B.A. from Dunbarton College of the Holy Cross.

Email: jcrussell@ufl.edu

Program Outline

DAY 1 Monday <i>Location: Piazzale Aldo Moro 7, Rome, Italy</i> Registration Desk Open 8:30-9:00 Coffee and Tea Service	DAY 2 Tuesday <i>Location: Piazzale Aldo Moro 7, Rome, Italy</i> Registration Desk Open 9:00-9:30 Coffee and Tea Service
Opening Session 9:00-10:30 <i>Welcome Address, Keynote Address, Opening Paper</i>	Poster Session, Sponsor Showcase 9:30-11:00 <i>Presentations in the digital poster corridor</i>
Morning Break 10:30-11:00	
Session One 11:00-12:30 <i>Exposing Grey Literature to Wider Audiences</i>	Panel Session 11:00-12:30 <i>Innovations in GL powered by Research Data</i>
Lunch 12:30-13:30	Lunch 12:30-13:30
Session Two 13:30-15:00 <i>Confronting Obstacles in Accessing Grey Literature</i>	Session Three 13:30-15:00 <i>Impact of Emerging Technologies & Social Media on Grey Literature</i>
Afternoon Break 15:00 -15:30	Pause 15:00-15:15
Introduction to Posters 15:30-16:30 <i>Short briefings by the Poster Presenters</i>	Closing Session 15:15-16:00 <i>Report by Moderators, Conference Handoff, Farewell</i>
 Main Conference Hall	 Digital Poster Corridor

Day One

Moderator: Margret Plank, German National Library of Science and Technology

09:00 – 10:30

OPENING SESSION

Welcome Address , Alberto De Rosa, National Research Council of Italy, CNR Central Library, Rome, Italy	14
Keynote Address , Imma Subirats, Food and Agriculture Organization of the United Nations, FAO	15
Opening Paper , Grey Literature and Research Assessment exercises: From the current criteria to the Open Science models Silvia Giannini, Rosaria Deluca, Anna Molino, and Stefania Biagioni, ISTI-CNR, Pisa, Italy	17

10.30 -11:00 Morning Break

11:00 – 12:30

SESSION ONE – EXPOSING GREY LITERATURE TO WIDER AUDIENCES

The Incorporation of Grey Literature into a newly formalized Research Services Program Charles Scott Dorris, Dahlgren Memorial Library; Georgetown University, United States	23
Indexing of Special Collections for Increased Accessibility Judith C. Russell, University of Florida, and Marjorie M.K. Hlava, Access Innovations, United States	24
Data Papers are Witness to Trusted Resources in Grey Literature: Driving Access to Data thru Public Awareness Dominic Farace and Jerry Frantzen, GreyNet, Netherlands; Plato L. Smith, University of Florida, United States	25
Public Access to the Dissertations in Russia Yuliya B. Balashova, Saint Petersburg State University, Russia	31

12:30 -13:30 Lunch

13:30 – 15:00

SESSION TWO: OVERCOMING OBSTACLES IN ACCESSING GREY LITERATURE

How open access policies affect access to GL in university digital repositories: A case study of iSchools Tomas A. Lipinski and Katie Chamberlain Kritikos, University of Wisconsin at Milwaukee, United States	37
Law, Liability, and Grey Literature: Resolving Issues of Law and Compliance Daniel C. Mack, University of Maryland, United States	43
Indexing grey multilingual literature in General Practice in the era of Semantic Web Marc Jamoulle and Marc Vanmeerbeek, University of Liège, Department of General Practice, Belgium Melissa P. Resnick, University of Texas, Health Science Center at Houston, USA Ashwin Ittoo, HEC Management School, University of Liège, Belgium Robert Vander Stichele, Heymans Institute of Pharmacology, University of Ghent, Belgium Elena Cardillo, Institute of Informatics and Telematics, National Research Council, Italy Julien Grosjean and Stefan Darmoni, D2IM, University of Rouen, France	49
Preserving and accessing content stored on USB-flash-drives: A TIB workflow Oleg Nekhayenko, German National Library of Science and Technology, TIB, Germany	55

15:00 - 15:30 Afternoon Break

15:30 – 16:30

INTRODUCTION TO CONFERENCE POSTERS AND SPONSOR SHOWCASE

On the afternoon of Day One, each person presenting a poster will have the opportunity to introduce their work in the Main Conference Hall. The Poster Session and Sponsor Showcase will continue on the morning of Day Two in the Digital Poster Corridor, where the presenters will be able to meet with conference delegates and participants in an informal setting. Those presenting during the conference Poster Session and Sponsor Showcase will have a chance to win the Poster Prize 2017.

Day Two

Moderator: Judith C. Russell, George A. Smathers Libraries, University of Florida

09:30 – 11:00

POSTER SESSION AND SPONSOR SHOWCASE

Digital Posters with Annotations	12
Other Invited Digital Posters	13

11:00 – 12:30

PANEL SESSION – INNOVATIONS IN GREY LITERATURE POWERED BY RESEARCH DATA

A Facet-based Open and Extensible Resource Model for Research Data Infrastructures Leonardo Candela, Luca Frosini, and Pasquale Pagano; ISTI-CNR, Italy	79
D4Humanities: Deposit of Dissertation Data in Social Sciences & Humanities – A Project in Digital Humanities Joachim Schöpfel, GERiiCO Laboratory, University of Lille; Hélène Prost, CNRS, associate GERiiCO Laboratory, France	85
Assessing the FAIRness of Data Sets in Trustworthy Data Repositories Peter Doorn, Data Archiving and Networked Services, DANS-KNAW, Netherlands	91
Assessing Data Management Needs and Practices to Enable Research Data Support Services Plato L. Smith and Jean Bossart, University of Florida Libraries, United States	97

12:30 -13:30 Lunch

13:30 – 15:00

SESSION THREE – IMPACT OF EMERGING TECHNOLOGIES & SOCIAL MEDIA ON GREY LITERATURE

Social Media Matters: Showing Up Online as Well as Ontime Julia M. Gelfand, University of California, Irvine Libraries, and Anthony Lin, Irvine Valley College Library, USA	103
Video is the new Grey Bastian Drees, Margret Plank, and Oleg Nekhayenko, National Library of Science and Technology, Germany	108
Impact of Emerging Information Technologies on Grey Literature Dobrica Savić, Nuclear Information Section, International Atomic Energy Agency, NIS-IAEA, Austria	109
Apps & Codes: Making profiles for fluid publishing contents Flavia Cancedda and Luisa De Biagi, National Research Council of Italy, CNR Central Library, Italy	115

15:00 – 15:15 Afternoon Break

15:15 – 16:00

CLOSING SESSION – REPORT MODERATORS, BEST POSTER, CONFERENCE HANDOFF, FAREWELL



Main Conference Hall



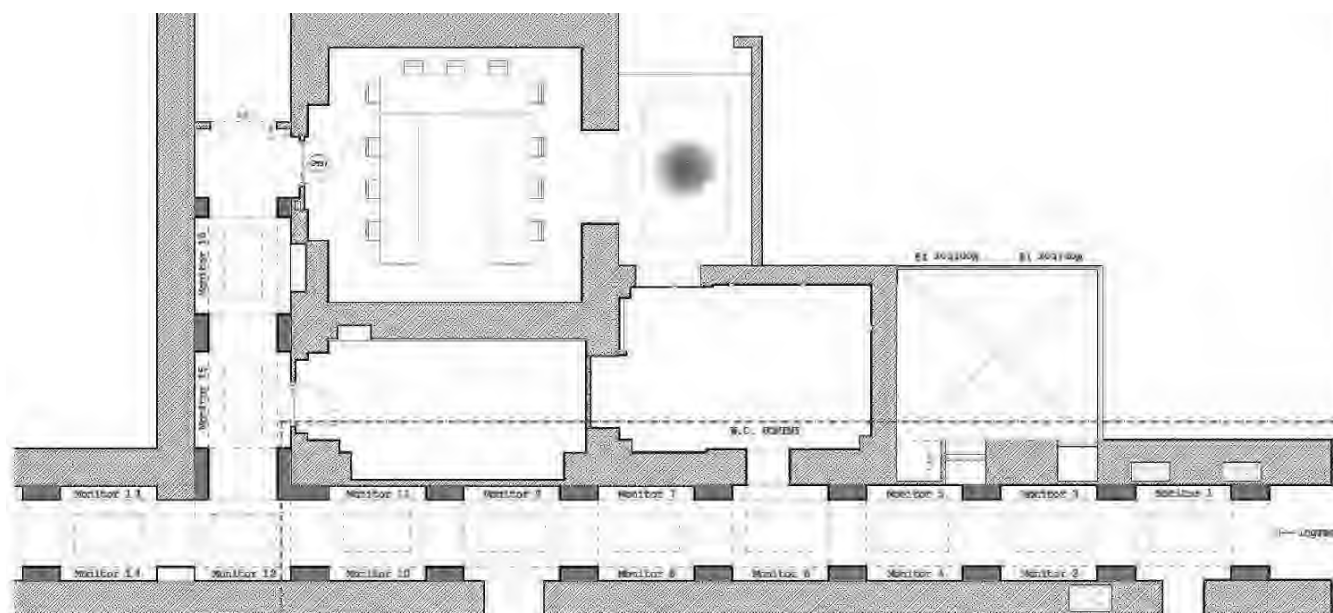
Digital Poster Corridor

Poster Session and Sponsor Showcase

Digital Posters with Annotations

Providing Access to Grey Literature: The CLARIN Infrastructure	60
<i>Sara Goggi, Gabriella Pardelli, Irene Russo, Roberto Bartolini, and Monica Monachini, CNR, Istituto di Linguistica Computazionale, "Antonio Zampolli", Italy</i>	
Data Visualization of a Grey Literature Community: A Cooperative Project	63
<i>Roberto Bartolini, Sara Goggi, Gabriella Pardelli, and Irene Russo, ILC-CNR Italy; Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands</i>	
Strategies for Teaching and Learning about Grey Literature, including the Dissemination and Exchange of Information	64
<i>Marcus Vaska, Alberta Health Services; Dean Giustini, University of British Columbia Library, Canada Hana Vyčítalová, National Library of Technology, Czech Republic</i>	
Collecting Grey Literature – Institutional Repository versus National Aggregator	69
<i>Petra Černohlávková and Hana Vyčítalová, NTK, National Library of Technology, Czech Republic</i>	
Grey Literature LibGuides or LibGuides about Grey Literature: A 2-Continent Environmental Scan of Common Themes & Trends	70
<i>Kathleen Noma Carlson, University of Arizona College of Medicine, USA; Joachim Schöpfel, Université de Lille Sciences Humaines et Sociales, France; Marcus Vaska, Alberta Health Services, Canada</i>	
Development of Visualization Search Service Model using Keyword Social Network for R&D Report	73
<i>YongJu Shin, Ki Seok Choi, ChulSu Lim, Jiseong Son, and Kwang-Nam Choi, NTIS Center, KISTI, Korea; Hwan Gue Cho, Department of Computer Science and Engineering, PUSAN National University, Korea</i>	
Classification Assistance for Korean R&D Project Using Machine Learning Techniques	75
<i>Jiseong Son, Kwang-Nam Choi, and Chul-Su Lim, NTIS Center, KISTI, Korea; Yu-Chul Jung, Department of Computer Engineering, Kumoh National Institute of Technology, Korea</i>	
GreyGuide Portal and Repositories "Sharing knowledge as early as possible"	77
<i>Stefania Biagioni and Carlo Carlesi, ISTI-CNR; Alessia Barsotti and Andrea Pardini, UNSC-ISTI-CNR, Italy</i>	

Floorplan Digital Poster Corridor



Poster Session and Sponsor Showcase

Other Invited Digital Posters

Video is the new Grey

Bastian Drees, Margret Plank, Oleg Nekhayenko, German National Library of Science and Technology, Germany

DANS-KNAW

Peter Doorn, Heidi Berkhout, and Harmen van der Meulen, Data Archiving and Networked Services, DANS Netherlands

INIS

Dobrica Savić, Nuclear Information Section; International Atomic Energy Agency, IAEA, Austria

Data Papers Project 2017 : GreyNet's Enhanced Publications 7 Years on

Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands; Plato L. Smith, University of Florida, USA

GOBI Library Solutions from EBSCO

Thomas Smith and Chris Mangione, EBSCO, USA

WorldWideScience.org : An International Partnership to Improve Access to STI and Research Data

Lorrie Johnson, WorldWide ScienceAlliance Operating Agent, USA

OpenAIRE : The OpenScience European Infrastructure

Michele Artini, Claudio Atzori, Miriam Baglioni, Alessia Bardi, Stefania Biagioni, Michele De Bonis, Andrea Dell'Amico, Sandro La Bruzzo, and Paolo Manghi, ISTI-CNR, Italy

The Grey Journal 2005-2017, Flagship Journal for the International Grey Literature Community

1982 2017 Grey Literature Network Service, GreyNet International

GreyNet Award Recipients 1999-2017 Recognition for Support and Achievements in the Field of Grey Literature

1982 2017 Grey Literature Network Service, GreyNet International

International Conference Series on Grey Literature 1993-2017 Vehicle for Extensive and Protracted Research in the Field of Grey Literature

1982 2017 Grey Literature Network Service, GreyNet International

The Grey Forum Series 2013-2017 Where Grey Literature provides Common Ground for Information Professionals

1982 2017 Grey Literature Network Service, GreyNet International

Welcome Address

Alberto De Rosa
Head Central Library
National Research Council of Italy



Alberto De Rosa is responsible for the National Research Council Central Library 'G. Marconi', the main Italian multidisciplinary library devoted to Science and Technics. He is involved in technological, administrative, and management activities related to scientific projects and Information Science in both National and European programs. He graduated in Economics at the Naples University 'Federico II'. He is a chartered accountant and statutory auditor (Register of the Italian Ministry of Economics and Finance). He obtained various Masters in Business and Management in the Public Administration and Research boards. From 1993 to 2013 he has been Administration Manager of CNR Research Institutes (e.g. the Institute of Biostructures and Bio-imaging); From 2002 to 2013 he has been Adjunct Professor of 'Company Structure' at the Naples University. And, from 2006 to 2014 he was Responsible for document management system, document flow and current archives at the CNR Institute of Biostructures and Bio-imaging. Email: alberto.derosa@cnr.it

Keynote Address

Facilitating Access to Global Grey Agricultural Literature

Imma Subirats,

Partnerships, Advocacy and Capacity Development Division;
Food and Agriculture Organization of the United Nations, OPCC-FAO-UN

Access to scientific research information is important for policy makers, governments and the scientific communities. In the agricultural domain, like many other disciplines, to access to latest scientific literature enables researchers to find solutions to local agricultural problems. This could be by discovering methods and practices that have worked elsewhere in disease control, management of pests and also improving agricultural productivity through implementing innovative approaches. And grey literature plays a key role in this context too.

In the agricultural domain, the Food and Agriculture Organization of the UN not only produces agricultural advice, but also curates and shares agricultural knowledge. For example through AGRIS, the International Information System for the Agricultural Science and Technology, a global database that provides access to bibliographic information on agriculture, science and technology. In September 2017, AGRIS reported more than 9 million records (9,025,192) and these come from 140 countries and contributed by 350 data providers. Of the 9 million records, 5.5 million are from journal articles; 34,866 from books; 287,724 from conference papers; 368,536 from bibliographies and 65,386 from Thesis.

This presentation will focus on the efforts to enable developing countries to gain access to publications, information and data in the fields of food, agriculture, environmental science and related social sciences by the Food and Agriculture Organization of the UN, and some additional projects and services to enhance the accessibility, visibility and usability of scientific research in agricultural sciences together with AGRIS.

Bionotes

Imma Subirats has worked as senior information management officer at the Food and Agriculture Organization of the United Nations (FAO) since 2006. She is the team leader of the Information and Data Access Services at the Partnerships, Advocacy and Capacity Development Division (OPC). With more than 20 years of experience in this sector, she is currently managing a large number of projects and services to enhance the accessibility, visibility and usability of scientific research in agricultural sciences. In this context, she is responsible of AGRIS, a free service with more than 8,500,000 records (1,500,000 to full text) to publications related to agricultural science and technology. She is also the AGORA Program Manager, program that delivers the most recent, high quality and relevant scientific literature to reduce the scientific knowledge gap and develop capacities to access and use the scientific literature in low-income countries. For the last 10 years, Ms Subirats has supported capacity development activities to enhance the creation, management, preservation and dissemination of information and data in agriculture. This includes to develop capacities to enable information and data managers to manage and exchange data efficiently, through online and face-2-face training programmes - in particular in Africa, Asia and Latin America. She is the co-chair of the Agricultural Data Interest Group (IDAG) of Research Data alliance (RDA), manager of the e-Agriculture community of practice and coordinator of the Agricultural Information Management Standards (AIMS) website. Email: imma.subirats@fao.org





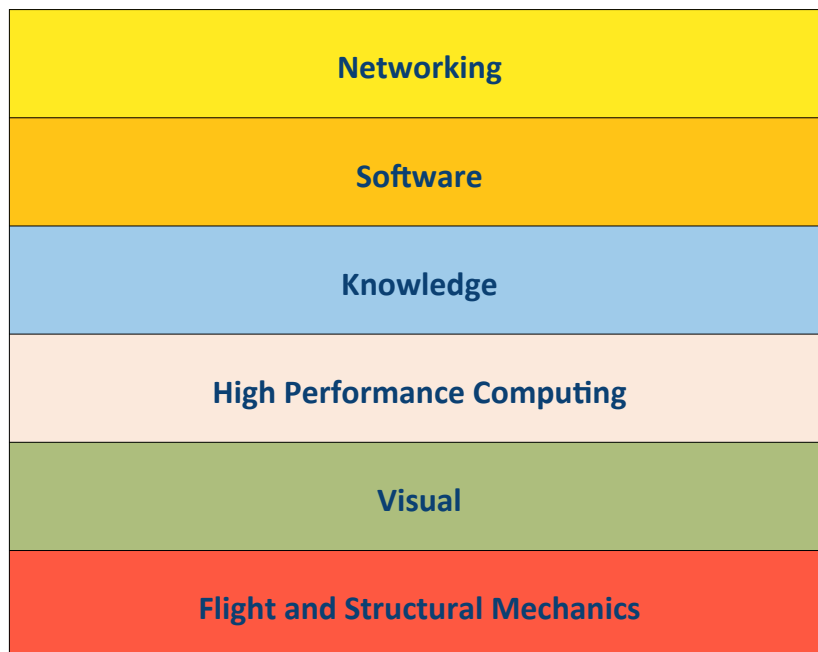
Institute of Information Science and Technologies “A. Faedo”

an Institute of the National Research Council of Italy CNR

ISTI is committed to produce scientific excellence and to play an active role in technology transfer.

The domain of competence covers Computer Science & Technologies and a wide range of applications.

The research and development activity of the Institute can be classified into 6 thematic areas



**CNR-ISTI, Via G. Moruzzi 1
56124 Pisa (PI), Italy
Area della Ricerca del CNR**

**Contact: +39 050 315 2403
segreteria scientifica@isti.cnr.it
<http://www.isti.cnr.it>**

Opening Paper

Grey Literature and Research Assessment exercises: From the current criteria to the Open Science models

Silvia Giannini, Rosaria Deluca, Anna Molino, Stefania Biagioni
CNR, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Italy

In the recent years the application of strategies, procedures and tools to evaluate the research have become subject of interest and their application is currently matter of discussion. The assessment exercises are regulated at national level and are carried out in different European countries such as France, United Kingdom and The Netherlands. In Italy the first research assessment exercise has been legislated in 2003 and entrusted to a specific Committee named Comitato di Indirizzo per la Valutazione della Ricerca (CIVR)¹. Three years later the CIVR and other committees have been replaced by a specific Agency named Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (ANVUR). This Agency, set up at the end of 2006, aims to «...rationalize the system of assessment of the quality of Universities, state and private Research Institutions beneficiary of public funds...» « The results of these activities managed by ANVUR represent a criteria to assign the state funds to Universities and Research Institutions».

It is not hard to imagine that the effects of this type of exercise has a strong political implication and determines a significant economic impact on the future of Universities and Research Institutions. The debate concerning the adopted methods and critical aspects about the assessment exercises is studied thoroughly at international level.

At the present time, ANVUR has completed two evaluation exercises of the quality of the research named Valutazione della Qualità della Ricerca (VQR): the first one spans the years 2004 - 2010; the second from 2011 to 2014. The work analyzes the environment of VQR in order to understand the organizational set-up, the operational models, the scientific areas involved in the process, the selection and evaluation criteria and indicators of the research products.

The work looks at the environment of VQR in order to understand the organizational set-up, the operational models, the scientific Areas involved in the process and the selection and evaluation criteria of the research products. More in detail, our work analyzes and compares the evaluation exercises conducted in Italy with the aim of verifying if and how Grey Literature is involved in the research evaluation processes. The article checked the types of products admissible for the research assessment and those actually presented by the researchers of Universities and Research Institutions. We measured the products from a quantitative point of view and observed their ramification in the different disciplinary fields rather than their transformation during the period of time taken into consideration. At the same time, we focused on the Open Science movement in order to understand what could be its role within the research assessment exercises and how it could affect the future of scholarly scientific communication.

¹The Committee analyzed the research products of 2001-2003 in order to evaluate the scientific performance of Universities, as well as state and private Research Institutions.

Bionotes

Silvia Giannini graduated and specialized in library sciences. Since 1987 she has been working in Pisa at the Institute for the Science and Technologies of Information "A. Faedo" of the Italian National Council of Research (ISTI-CNR) as a librarian. She is a member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She is responsible of the library automation software "Libero" in use at the CNR Research Area in Pisa and coordinates the bibliographic and managing activities of the ISTI library team. She cooperates in the design and development of the PUMA (Publication Management) & MetaPub, an infrastructure software for institutional and thematic Open Access repositories of published and grey literature produced by CNR. Email: silvia.giannini@isti.cnr.it



Bionotes *CONTINUED*

Anna Molino graduated in Linguistics at the University of Pisa in 2010. Since 2013, she works at CNR - ISTI ("Istituto di Scienza e Tecnologie dell'Informazione - A. Faedo") as member of the Networked Multimedia Information Systems Lab. (NeMIS). She has worked as project assistant and financial manager in various EU funded and national research projects for the Digital Libraries group of the NeMIS lab. She contributes in the language revision and translation of scientific papers. Email: anna.molino@isti.cnr.it



Stefania Biagioni graduated in Italian Language and Literature at the University of Pisa and specialized in Data Processing and DBMS. She is currently a member of the research staff at the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), an institute of the Italian National Research Council (CNR) located in Pisa. She is head librarian of the Multidisciplinary Library of the CNR Campus in Pisa and member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She has been the responsible of ERCIM Technical Reference Digital Library (ETRDL) and currently of the PUMA (PUBlication MAnagement) & MetaPub, a service oriented and user focused infrastructure for institutional and thematic Open Access repositories looking at the DRIVER/OpenAire vision, <http://puma.isti.cnr.it>. She has coauthored a number of publications dealing with digital libraries. Her activities include integration of grey literature into library collections and web access to the library's digital resources, including electronic journals and databases. She is a member of GreyNet since 2005. Since 2013 she is involved on the GreyGuide Project. Email: stefania.biagioni@isti.cnr.it



Nineteenth International Conference on Grey Literature. Public Awareness and Access to Grey Literature
October, 23-24, 2023 - Rome, CNR, Italy

**Grey Literature and Research Assessment exercises:
from the current criteria to the Open Science models**

Silvia Giannini, Rosaria Deluca, Anna Molino, Stefania Biagioni
CNR-ISTI, Pisa, Italy

Summary

- Scenario & Objectives
- The VQR Environment
 - ✓ Organization and Methods
 - ✓ The VQRs "objects"
- Tracking the Grey Literature
- Analysis of data and results
- Open Science and Grey Literature
- Conclusions

Scenario and Objectives

- In the recent years the application of strategies, procedures and tools to evaluate the work of researchers have become subject of interest and their application is currently matter of discussion.
- The assessment exercises are regulated at national level and carried out in different European countries, e.g. France, United Kingdom and the Netherlands.
- At the present time, two evaluation exercises of the quality of the research named *Evaluation of Research Quality (VQR)* were realized in Italy. The first one spans the years 2004 – 2010 (VQR1); the second from 2011 to 2014 (VQR2).
- The conceptual challenges taken on by the Open Science (OS) movement may be crucial for the evolution of these matters.

The work analyzes the VQRs objects and methods with the aim of verifying:

- *if and how* Grey Literature is involved in the research evaluation processes;
- *what will be the future* of the scholarly scientific communication according to the instances of the Open Science movement.

Organization and Methods: the VQRs environment

What is evaluated?

The research conducted in both state and private universities, public research bodies and other public and private subjects funded by the government

- ✓ Researchers, assistant professors, associate professors, full professors...

Who are the evaluators?

National Agency for the Evaluation of Universities and Research Institutes (ANVUR)



Groups of experts for the evaluation – GEV

The methodology

- Qualitative assessment
- Quantitative assessment
- Informed peer review

VQR 2004-2005		VQR 2011-2015	
Class of merit	Score	Class of merit	Score
A. Excellent	1	A. Excellent	1
B. High-level	0.8	B. High-level	0.7
C. Fair	0.5	C. Fair	0.4
D. Acceptable	0	D. Acceptable	0.3
E. Limited	-1	E. Limited	0
F. Not evaluable	-2	F. Not evaluable	0

ANVUR - Quality steps

Organization and Methods: the VQRs environment

Taxonomy based on macro disciplinary areas and Scientific Disciplinary Sectors (SSD)

Area	Description
1	Computer science and Mathematics
2	Physics
3	Chemistry
4	Earth sciences
5	Biology
6	Medicine
7	Agricultural and veterinary sciences
8	Civil engineering and Architecture
9	Industrial and computer engineering
10	Artistic sciences
11	Historical, philological, psychological and pedagogical sciences
12	Legal sciences
13	Economics and Statistics sciences
14	Social and political sciences

Area	Description
1	Computer science and Mathematics
2	Physics
3	Chemistry
4	Earth sciences
5	Biology
6	Medicine
7	Agricultural and veterinary sciences
8a	Architecture
8b	Civil engineering
9	Industrial and computer engineering
10	Artistic, philological, literary and historical sciences
11	Quantitative
12a	Historical, philological and pedagogical sciences
12b	Psychology
12c	Legal science
13	Economics and Statistics sciences
14	Social and political sciences

It's clear the substantial overlapping between macro-areas in VQR1 and VQR2.

Scientific Disciplinary Sectors

Area 1

Computer science

Logic

Algebra

Geometry

Complementary maths

Mathematical analysis

Area 2

Experimental physics

Theoric physics, mathematical

models and methods

Physics of matter

Nuclear and subnuclear physics

Astronomy and Astrophysics

Physics for the Earth system

Area 3 ...

Organization and Methods: the VQRs "objects"

Documentary categories in VQR1 (2004 – 2010)

- Papers in journals
- Books, chapters of books, and conference proceedings provided with ISBN
- Critical editions, translations, and scientific comments
- Patents
- Compositions, drawings, design, performances, exhibitions and organized expositions, artwork, prototypes, artworks and related designs, databases and software, thematic maps

Not admissible products in VQR1

- Editorial and curatorial activities
- Conference abstracts (even if published in journals)
- Texts or software used for educational and dissemination purpose only
- Routine or laboratory tests
- Internal technical reports

Organization and Methods: the VQRs “objects”

Documentary categories in VQR2 (2011 – 2014)

Scientific monographs

Research monograph, Concordance, Scientific comment, Annotated bibliography, Critical editions of texts, Critical editions of excavations, Publication of unedited sources, Critical manuals (not for educational purpose only), Grammar and science dictionaries, Translations of books (upon GEV's decision).

Articles in journals

Scientific paper, Review essays, Letters, Contribution to a Forum upon invitation of the editorial staff, Case notes, Translations in journal.

Contributions to books

Scientific articles in peer-reviewed conference proceedings, Foreword and afterword in the form of essay, Curatorship of books with introductory essay, Catalogues with introductory essay, Critical entries in dictionaries or encyclopedias, Translations in book (upon GEV's decision), Catalographic records, bibliography or corpora.

Other types of scientific products

Compositions, Drawings, Architectonic projects, Performances, Exhibitions, Prototypes of art and related projects/designs, Database and software, Thematic maps, Psychological evaluation, Audiovisual material.

Patents

The category Patents is always considered as evaluable, but it may be attributed to class A or B only if internationally renowned or licensed.

Not admissible products in VQR2

- *Manuals and texts for educational purpose only*
- *Review of a single article not showing any critical analysis of the literature on the topic*
- *Short, non-original encyclopedia or dictionary entries*
- *Short, non-original case notes*
- *Short catalographic records*

Tracking the Grey Literature

The process of identification of the Grey Literature inside the various documentary typologies was based on the following considerations:

- the evaluation exercises mainly founded their bibliometric analysis on the contents of the two commercial databases Web of Science and Scopus;
- the majority of the literature indexed by the two databases is published by commercial publishing companies;
- the use of the databases restricts the contents only to the references indexed (based on ownership criteria);
- the algorithm for the assessment of *papers in journals* takes into account the number of citations of a paper and the corresponding bibliometric indicator of the journal;
- the GEV's criteria specify that products listed in *Other types of scientific products* are evaluated making reference to their characteristics, not to their formal publication.

With specific reference to what listed above we agreed on ascribing some groups of products to the *non-conventional literature*.

Analysis of data and results

In both exercises and for each disciplinary area, the most significant numbers are referred to the entries *papers in journals*, *papers in books* e *papers in proceedings*. The *papers in journals* still represents the more widely evaluated category.

The frequency of the Grey Literature is 0.61% in VQR1 and 0.74% in VQR2. 

✓ In VQR1 the most relevant percentages are those referred to the Areas 7 - Agricultural and veterinary science, 8 - Civil engineering and architecture, and 10 - Antiquity, philological-literary and historical and artistic sciences.

✓ In VQR2 the Areas 8a - Architecture (mainly in 2011) and 12 - Legal science (especially in 2011 and 2014) gather the majority of the grey products.

The extension of the documentary categories in VQR2 influenced the incidence of the Grey categories in VQR2, where we find products that did not appear in the previous evaluation.

Some of the products present in both evaluations do not show significant annual variations, with the only exception of some categories such as *entries* as well as *exhibition* and *prototype*.

Open Science and Grey Literature...a perfect marriage..

The combination between the principles and the tools of the Open Science may represent a favorite dissemination channel for the Grey Literature, which conversely may become the primary source of the Open Science.

- The OS meets the demand of up-to-date, faster, more effective and less expensive dissemination modalities, making accessible the documentary typologies currently not publicized and unavailable.
- The OS meets the demand of new criteria for the research assessment through the use of different or complementary measures. It revolutionizes the modalities for the scientific foundation and dissemination, as well as the objectives, the models and the assessment methods.
- The OS highlights the need of making a better use of technology to share the outcomes and connect data and publications in a more effective way.
- The OS guarantees that the management and the curation of the data is attributed to the authors and not to the publishers.

The Grey Literature community formulated The Policy Development for Grey Literature Resources looking at the new principles of Open Science, and joining the new ways of creating, sharing and evaluating science (Pisa Declaration 2014).

Conclusions

The Italian Research Assessment Processes do not completely exclude Grey Literature. However, they are almost exclusively based on the analysis of commercially distributed products.

This is due to:

- The non-eligibility of some research products (e.g. *educational material, technical reports, commentary, obituary, erratum...*).
- The lack of interest in entries such as *preliminary studies, progress reports, accounts, search results, dossier, market researches, normative documents, feasibility studies*, etc..
- The disadvantage in the assessment of scientific products other than articles in journals.
- The impact of the evaluation criteria on researchers leads to the philosophy of *Public or Perish*: the researchers publish only scientific articles in prestigious journals.

The risk is to produce *fashionable* research rather than *quality* research.

Conclusions

- The Open Science models launch new criteria making visible, recognizable, identifiable and usable otherwise unavailable documentation and makes it possible to evaluate non-traditional materials.
- In an increasingly connected world, it is time for the scientific institutions and politics to interrogate, exchange experiences, build networks across national borders, allowing the growth of a new dialogue between science and society. Europe plays a key role in integrating scientific activity into the social, cultural and political economic landscape.
- Cultural, political and economic changes are necessary in order to realize the Open Science and support a wider view of the design of the research, the management of the projects, and for the dissemination of the results.

We think, in this new ecosystem of the scientific communication, that the Grey Literature can be exposed to a wider audience, can remove the obstacles in accessing and can enhance its contents and visibility thanks to the advanced technologies and the use of social media.

The Incorporation of Grey Literature into a newly formalized Research Services Program

Charles Scott Dorris, Dahlgren Memorial Library,
Georgetown University, United States

The ultimate goal of this project is to increase knowledge and use of grey literature resources amongst the librarians and patrons of a health sciences library at an academic medical center.

Up until now, librarians at Dahlgren Memorial Library (DML) have informally and inconsistently introduced the concept of grey literature to the students, faculty, and researchers at Georgetown University Medical Campus and to the clinicians and nurses at MedStar Georgetown University Hospital. The topic of grey literature may be tangentially referenced to in research consultations or workshops, and our online resource guide on this topic has yet to receive steady traffic. Moreover, we do not have marketing materials in place to help promote the use of grey literature, nor a system to track our interactions when we discuss it with our patrons. In sum, we have had a passive approach regarding the teaching of grey literature and we want to transition to a proactive approach by officially incorporating it into our services and our librarians' repertoire of knowledge.

This year, through a newly redesigned librarian position, we are formalizing a new research services program that will highlight different ways librarians can promote usage of grey literature, primarily during the initial stages of the research life cycle. Beginning with a background literature search and an environmental scan of all Association of American Medical Colleges library websites, we seek to learn how other health sciences libraries promote grey literature and whether it is incorporated into their research and education services.

We have sought out webinars that focus on grey literature to build our foundation and have attended multidisciplinary conferences such as GL18 in New York City. To help realize our goal, we will offer a series of continuing education workshops for librarians on grey literature, both the topic and the means to locate it. We will introduce a research worksheet that contains a dedicated section for grey literature so librarians and patrons are consistently prompted to consider the use of these resources. In order to help promote and market grey literature, we will reach out to departments and committees to raise awareness of how our librarians can assist with their research goals. We wish to build upon our already strong integration in some academic programs such as the Medical School's Medical Education Research Scholars track and the Department of Physiology's Complementary and Alternative Medicine program. Lastly, we intend to capture librarian usage of grey literature resources and statistics using LibInsight.

Bionote

Charles Scott Dorris is the Research Services Coordinator at Dahlgren Memorial Library at Georgetown University Medical Center. He has been in this role since January 2017 after 8.5 years as the Digital Information Services Librarian. In this relatively new role, he creates, coordinates, and markets initiatives to support research on the medical campus. He also provides library instruction support for the affiliates at MedStar Georgetown University Hospital. He holds a Master's in Library and Information Science from the University of Pittsburgh and a Bachelors of Arts in Anthropology from The Pennsylvania State University. Email: csd24@georgetown.edu



Indexing of Special Collections for Increased Accessibility

Judith C. Russell, University of Florida,
Marjorie M.K. Hlava, Access Innovations, United States

Challenge of Discovery: Recent large-scale initiatives focused the attention of the Smathers Libraries at the University of Florida on the need for significantly expanded and enhanced metadata for our digital collections, both retrospective and prospective. This requires new tools and changing roles and responsibilities for cataloging/metadata staff, including the application of automated processes, Improved and consistent metadata practices, and the development of new taxonomies. Projects that are described include the new genealogical initiatives with Internet Archive and Family Search, Portal of Florida History, the Digital Library of the Caribbean (dLOC) and the Cuban Heritage Initiatives.

We have concluded the pilot project on automated indexing and metadata generation for the Portal of Florida History using 25,000 full text records and the JSTOR thesaurus instead of the Library of Congress subject headings to automatically index the collection and create a search portal. This paper gives a report on the results of the pilot and early application of the process to the entire UF Digital Collections to extract information from our complete digital collections (over 600,000 items/12 million pages) for the Portal.

Results of the pilot show significantly increased retrievability, greater depth of accessibility via detailed subject metadata and then explores application to the entire digital collection. The idea of a more automated processes going forward to allow traditional cataloging to focus on the things that need individual attention and use automated tools to develop and improve metadata for other materials is explored since the large number of items (over 12 million and growing at 100,000 pages per month) makes it impractical to use traditional means. We are working with tools that have been developed for information products and services, but can be applied effectively to library collections.

This paper covers three of the main themes of the conference: 1) Exposing Grey Literature to Wider Audiences using an open search portal 2) Confronting Obstacles in Accessing Grey Literature through limited metadata tagging and keeping the collection access obscure as sub sets of the Library OPAC, and 3) the Digital Preservation project covering everything from old newspapers, personal papers and county historical collections is the Lifeline for Grey Resources but if not widely available through the internet and deeply tagged using a controlled vocabulary the simple scanning of papers only creates the new microfilm dilemma of locking data away in inaccessible places.

Data Papers are Witness to Trusted Resources in Grey Literature: Driving Access to Data thru Public Awareness

Dominic Farace and Jerry Frantzen, GreyNet, Netherlands;
Plato L. Smith, University of Florida, USA

In 2011, GreyNet embarked on an Enhanced Publications Project with DANS, Data Archiving and Networked Services in an effort to circumvent the data versus document camps entrenched in grey literature communities¹. The results of that two year project served to incorporate in GreyNet's workflow the acquisition, indexing, and linking of research data with their full-text and accompanying metadata^{2,3}.

Recently, GreyNet discussed with a data management librarian from UFL, University of Florida a proposed follow-up project dealing with data papers – a new document type within grey literature. Data papers are defined as scholarly publications of a searchable metadata document describing a particular online accessible dataset or a group of datasets published in accordance to standard academic practices. As such, data papers represent a scholarly communication approach to data sharing.⁴

The outcome of that discussion has led to the formation of a project team with the twofold purpose of producing and publishing a set of data papers originating in the field of grey literature, and in so doing raise awareness to this new document type by demonstrating its value for library and information science.

The method of approach includes the construction of an online standardized template that encompasses a data paper, defining the population asked to complete the template, instruction and further contact with the authors/researchers during the course of the project, along with an analysis of the results. While there are no direct costs associated with this project, each of the partners is committed to allocate human and material resources needed to carry out their related tasks.

The anticipated outcome of the project would provide a tested template that could be used by other grey literature communities in the production of data papers. It would demonstrate how OA, DSA⁵ and FAIR⁶ principles are implemented and reinforced via data papers. And, it would further provide examples of how data citations can generate trusted bibliographic references.

1 Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012 <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:53456>

2 GreyNet's Enhanced Publications Project: Tracking and Backtracking Data, 2012
<http://greyguide.isti.cnr.it/index.php/49-gl14/gl14-slide-share/416-gl14-farace-et-al-2>

3 Frequently Asked Questions (FAQ): Enhanced Publications Project (EPP), 2013
<http://www.greynet.org/images/FAQ-EPP.pdf>

4 Data Papers https://en.wikipedia.org/wiki/Data_publishing#Data_papers

5 DSA, Data Seal of Approval <https://www.datasealofapproval.org/en/>

6 FAIR Data Principles <https://www.force11.org/group/fairgroup/fairprinciples>

Bionotes

Dominic Farace is Head of GreyNet International and Director of TextRelease, an independent information bureau specializing in grey literature and networked information. He holds degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked since 1976. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (SWIDOC/KNAW), Farace founded GreyNet, Grey Literature Network Service in 1992. He has since been responsible for the International Conference Series on Grey Literature (1993-2013). In this capacity, he also serves as Program and Conference Director as well as managing editor of the Conference Proceedings. He is editor of The Grey Journal and provides workshops and training in the field of grey literature. Email: info@greynet.org



Jerry Frantzen graduated in 1999 from the Amsterdam University of Applied Sciences/Hogeschool van Amsterdam (HvA) in Library and Information Science. Frantzen is the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet, Grey Literature Network Service, as a freelance technical consultant. Email: info@greynet.org



Plato Smith is the Data Management Librarian at the University of Florida with experience in academic research libraries, digital libraries, and data management. He received his doctorate in the field of Information Science from the School of Information within the College of Communication and Information at Florida State University, Florida's iSchool, Summer 2014. From 2005 to 2012, he was Department Head for the FSU Libraries' Digital Library where he developed, populated, and managed digital collections in the FSU Libraries' digital content management system, DigiNole Repository, and electronic theses and dissertations (ETDs) institutional repository. Email: plato.smith@ufl.edu



**Data Papers are Witness to Trusted Resources in Grey Literature:
Driving Access to Data thru Public Awareness**

Dominic Farace and Jerry Frantzen
GreyNet International, Netherlands



Plato L. Smith II
University of Florida Libraries, United States



Nineteenth International Conference on Grey Literature
National Research Council of Italy, Rome – October 23-24, 2017

Definition | Data Paper

Scholarly publications of a searchable metadata document describing a particular online accessible dataset or a group of datasets published in accordance to standard academic practices. As such, data papers represent a scholarly communication approach to data sharing ([wikipedia.org](https://en.wikipedia.org/wiki/Data_paper)).



Background | Data Paper Project

Enhanced publications "A publication that is enhanced with three categories of information: research data, extra materials, and post-publication data". ([DRIVER-II](#))

Enhanced publications combine textual resources *i.e.* documents intended to be read by human beings, which contain an interpretation or analysis of primary data.

Enhanced publications inherently contribute to the review process of grey literature as well as the replication of research and improved visibility of research results in the scholarly communication chain.

R&D | Data Collection

2011

Survey

Author/
Researcher
'Attitudes on
Sharing Data'

**2012 ...**

Linking Data

Data Acquisition,
Published Datasets,
Workflow, and
Linking Data

**2017**

Data Paper

Template,
Acquisition,
Publication,
and Training



Template | Data Paper

The template for this project is based on a compilation derived from 3 existing templates: RDJ, JOHD, and JOPD

The template consists of five main sections:

- 1 Overview
- 2 Methods
- 3 Dataset Description
- 4 Potential Reuse
- 5 References



Population | Data Papers Project



In 2012, GreyNet's 1st dataset was entered in the DANS Archive

GreyNet now has a collection of **30 datasets** in

Review Process | Data Papers

1. Submissions are first checked against the standardized template. The author(s) may be asked to provide additional information.
2. The (revised) data paper is then sent to the Data Management Librarian for review.
3. Once reviewed the data paper is entered in the DANS Archive as a preprint, where it is CC0 licensed and shares in the DSA, Data Seal of Approval.
4. As such the Data Paper implements FAIR Principles: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable

Publication | Data Papers

The data paper is first published as a **preprint** alongside its corresponding dataset in the DANS Archive. There it is preserved and remains openly accessible via a license.



The data paper is then scheduled for republication in TGJ, The Grey Journal (ISSN 1574-1796), where it receives coverage through multiple A&I Services and is available in full-text via the EBSCO LISTA-FT database.



Results | Data Papers Project

Since the initial request to **15** first authors responsible for GreyNet's Collection of published datasets in the DANS Archive, results to date show:

- The Data Paper Template complied and edited for this Project has proven to be efficient.
- **Six** authors have indicated an interest in drafting a Data Paper, **five** of which are now published in DANS and **two** in TGJ.
- Data Papers are now listed as a new GL Document Type appearing in the **GreySource** index.
- Data Papers lend added support to the **Pisa Declaration on Policy Development for Grey Literature Resources**.



Spin-Off | Data Papers Project

The Grey Journal

Authors, who have published in The Grey Journal outside of the GL-Conference Series and whose data is not in the DANS Archive, have indicated a willingness to submit their data and complete a Data Paper.



GreyForum Series

"Data Papers, A Trusted Tool in Research and Data Sharing"

This workshop and training is being developed and will be offered to information professionals. The emphasis lies in raising awareness to data access and its potential reuse for science and society via Data Papers.



Follow Up | Data Papers Project



GreyNet
Workflow



User
Statistics



Data Citations
and References



Author
Feedback

On behalf of the Project Team



Thank You

And, we hope to hear from you soon!

info@greynet.org

GreyNet

Grey Literature Network Service

Public Access to the Dissertations in Russia

Yuliya B. Balashova

Saint Petersburg State University, Russia

In the modern Russia as a part of the global world, scientific communications are developing quite actively. Their growth is caused, among other things, by the task of wide audience familiarizing with the scientific thought achievements. New popular science media are being created, as well as aggregators of scientific news, scientific festivals are held, science promoted through the social networks. However, in the general scientific context, status of dissertations (thesis) themselves remains not completely clear. Accordingly, the public discussion on these issues seems problematic.

This state of affairs is due primarily to the publishing status of the dissertations, as well as the author's abstracts to them. They are published "on the rights of the manuscript", in other words, they are not formally copyright subject. In addition, in order to have a scientific career, the Russian scientist has to defend not one, but two dissertations. The first thesis – "candidate", corresponding to the accepted Ph.D., and the second thesis defended for a doctoral degree and correspond to the title existing in a number of countries – "habilitated doctor". Thus, scientists in Russia write, as a rule, two dissertations and two abstracts. At present, a special process for acquainting the scientific community with the research results has been adopted. For this, several months before the defense, the author's thesis and author's abstract are published: 1) on the website of the organization where the work is prepared, 2) on the website of the "Higher Attestation Commission" of the Russian Federation. Reviews for the thesis are also published on the website of the organization where the thesis is being defended. The defense itself has a public character in the way of broadcast live. After defense, the volume of the dissertation is sent to the Russian State Library, the main scientific library in the country. There it is stored both in paper and electronic form. The author's abstracts of dissertations are also sent to the main scientific libraries of Russia. This information is open and accessible to all interested.

But the mass audience is mainly acquainted with scientific theses in a specific way: mostly on the path of scandals, covered in the media. The main source of exposing falsified dissertations is the network resource "Dissernet". It was thanks to its efforts that a scandal with the thesis of the Minister of Culture V. Medinsky received considerable coverage.

Thus, both the mass and specialized audience for the first time received such wide access to the theses. The next urgent task is to integrate the achieved result into the international scientific community.

Note: The study is supported by the Russian Humanitarian Foundation (grant 16-03-50128). The author specializes in theoretical understanding and practical implementation issues relating to the interaction between science and society.

Bionote

Yuliya B. Balashova, is Doctor of Philology, journalist, Associate Professor at St Petersburg State University, Russia. Email: j.balashova@spbu.ru





Public Access to the Dissertations in Russia

Prof. Yuliya Balashova
St. Petersburg State University, Russia

j.balashova@spbu.ru
ubalash@gmail.com



System of Academic Degrees Awarding

Western

(mostly one-level system of
scientists degrees)

1) Ph.D.



Russian

(two-level system of
scientists degrees)

1) Candidate of sciences

2) In some European countries –
habilitated doctor:

Doctor habilitatus, Dr. habil.



2) Doctor of sciences:
Dr. habil.,
Doctor of Sciences



Correlation between university autonomy and state regulation in awarding academic degrees

*Who could accredited doctoral (PhD) programs,
carry out quality assessment:*

- 1) State, authorized with the state structures: France, Norway, Spain, Sweden, United Kingdom, New Zealand
- 2) Socio-professional organizations: USA, Canada
- 3) Academy of Sciences (Netherlands)

Who has the right to assign academic degrees:

- 1) State (Russia, post-Soviet countries)
- 2) Universities (the rest of the world)

Obligatory stages of the dissertations publicity before defense

- 1) On the website of the organization, where the work was prepared. There are published the full text of the dissertation, the author's abstract of the dissertation, information about official opponents, and the leading organization, preparing review for the dissertation, as well as the place and time of dissertation's defense. From 2012 at Saint Petersburg State University all dissertations defenses are accompanied by online translation on the university website (<https://disser.spbu.ru/ph-d-spbu.html>)
- 2) High Attestation Commission under the Ministry of Education and Science of the Russian Federation. Since 2006 H.A.C. publishes on it's website the same information, only in a reduced amount, and with reference to the organization's website (<http://vak.ed.gov.ru>)

The main results of the dissertation research should be published in the form of articles in the list of peer-reviewed scientific publications.
High Attestation Commission of the RF on a regular basis revises this list



Obligatory stages of the dissertations publicity after defense

- 1) Obligatory dispatch of the dissertation abstract to the leading libraries of Russia, and also National Library of Belarus (<http://www.nlb.by>)
- 2) The full written text of the dissertation is sent to the Russian State Library. The main library of the country provides public access to the digitized abstracts, and thesis themselves. Site navigation is available in English (<http://www.rsl.ru>)
- 3) A compulsory copy of both the thesis and the author's abstract is sent by mail to the Russian Book Chamber (<http://www.bookchamber.ru>). In accordance with law, a certain number of copies of any printed publication (books, brochures, periodicals, dissertations, musical notes, geographic maps and atlases) must be sent to the Russian Book Chamber

Other (additional) online channels for dissertations distribution

For example: 'Library of Dissertations' (DsLib.net). They provide both free and paid access to the thesis text, publish information about upcoming defenses. At the same time, monetary deductions to authors of dissertations are declared

Another example: 'CyberLeninka' (cyberleninka.ru) positions itself as a scientific open access electronic library, the main tasks of which are popularization of science and scientific activity, public control of the quality of scientific publications, the development of interdisciplinary research, the modern institute of scientific review, and the increasing citation of Russian science. 'CyberLeninka' is built on the basis of the open science paradigm

The fundamental problem is that thesis and abstract both are published "on the rights of the manuscript", in other words, they are not formally copyright subject. So, private distributors don't pay in fact any royalty, that violates the norms of morality, but not the law

The network encyclopedia 'Famous Scientists' (biographical data of scientists) is a project of the Russian Academy of Natural History. It is also published the main results of the dissertations research



The main source of exposing falsified dissertations is network 'Dissernet'





'Dissernet' – "a free network of experts, researchers and reporters, who dedicate their work to exposing scammers, falsifiers and liars. The participants of the community are cooperating with the joint efforts based on the principles of the network distribution of labor and using modern computer technologies, to counteract illegal frauds and infringements in the field of scientific and educational activities, especially in the process of defending dissertations and appropriating academic degrees in Russia"

Statistics. More than 10,000 dissertations have been analyzed since 2013. Based on 'Dissernet' requests 'High Attestation Commission' made a decision on deprivation of academic degrees



In the present time the procedure for awarding academic degrees begins to change.

A number of leading Russian universities and scientific organizations will have the right to self-award scientific degrees. On the one hand, thus, Russian universities become more integrated into the Western system. the supervising role of the state is abolished, the dissertational councils do not act on a permanent basis, but are created in accordance with the subject matter of the concrete work, the defenses take place in English. But such innovations cause a lot of criticism.

It is believed that in a democratic society decentralization is extremely constructive. In the scientific policy of modern Russia, this kind of decentralization causes great concern in terms of maintaining a high level of publicity.

One of the most serious concerns: will not the theses be less public, and the thesis defense process more closed?



Conclusion

- 1) Never before science in Russia has been so public. This result was achieved due to the efforts both of the state (scientific policy), and society
- 2) The next step is to integrate this openness into the Western system



FEDLINK

an organization
of federal agencies
working together
to achieve optimum use of
resources and facilities
of federal libraries
and information centers
by promoting
common services,
coordinating and sharing
available resources, and
providing continuing
professional education.



101 Independence Ave, SE ~ Washington, DC 20540-4935
FEDLINK Main Number (202) 707-4800
FEDLINK Hotline (202) 707-4900

Strategic Sourcing



Currently, more than 20 federal agencies, both military and civilian, including FEDLINK, participate in the Federal Strategic Sourcing Initiative (FSSI). FSSI was created in 2005 by the Dept. of the Treasury, the Office of Management and Budget, and the General Services Administration to identify products and services that can be purchased more efficiently and to provide centralized acquisition functions for a variety of products to streamline efficiency and to reduce costs to the federal government.

How open access policies affect access to grey literature in university digital repositories: A case study of iSchools

Dr. Tomas A. Lipinski and Katie Chamberlain Kritikos

School of Information Studies, University of Wisconsin-Milwaukee, United States

Problem/Goal: An issue of interest to library and information science (“LIS”) scholars and practitioners is how open access policies can affect the access and use of grey literature in university digital repositories. Open access (“OA”) refers to research placed online free from all price barriers and from most permission barriers (Suber, 2015), allowing unfettered access to scholarship and promoting open scholarly communication (Banach, 2011; Eysenbach, 2006). OA may apply research published traditionally, such as books (Schwartz, 2012) and academic articles (Suber, 2015), and published non-traditionally (*i.e.*, grey literature), such as student electronic theses and dissertations (Schöpfel & Prost, 2013; Schöpfel & Lipinski, 2012).

The treatment of grey literature in university digital repositories is of particular import due to “the ephemeral and changing nature of grey publication types, editions, and formats” (Rucinski, 2015, p. 548; see Farace & Schöpfel, 2010). The access and use of grey literature in these repositories is often executed through an OA policy. There is a gap in the literature, however, regarding best practices for drafting and implementing OA policies that promote unfettered access to grey literature.

Research Method/Procedure: This paper analyzes OA policies from a sample of U.S. iSchools, created by cross-referencing the iSchool Directory (iSchools, 2014) with the top twenty-five best LIS programs ranked by U.S. News and World Reports (U.S. News, 2017). Initial analysis shows that of the twenty-two iSchools in the sample, all schools have university digital repositories but only fifteen have OA policies. The project maps these policies against variables drawn from the benchmark for open scholarly communication, the Harvard Open Access Project’s *Good Practices for University Open-Access Policies* (Shieber & Suber, 2017; 2015; 2013; see also Harvard OSC, 2015; Nguyen, 2008).

Results: The goal of this paper is to understand how OA policies at university digital repositories affects access to grey literature in an ever-changing information landscape. Based on the analysis of the sampled iSchool OA policies and the Harvard OA policy variables, it recommends best practices for drafting and implementing OA policies that provide unfettered access to grey literature in university digital repositories.

Keywords: grey literature, open access, policy, access, university repository, best practice, scholarly communication, library science, information science

References

- Banach, M. (2011). *The benefits of managing and publishing ETDs “in house” using an open access repository*. Presentation at USETDA 2011, Orlando, Florida, May 18-20, 2011.
- Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, 4(5), e157, 0692-0698. doi:10.1371/journal.pbio.0040157.
- Farace, D. J. & Schöpfel, J. (2010). *Grey literature in library and information studies*. Berlin: DeGruyter Saur.
- Harvard Office for Scholarly Communication [Harvard OSC] (2015). Model open access policy. *Harvard Library Office for Scholarly Communication*. Retrieved from <https://osc.hul.harvard.edu/modelpolicy/>.
- iSchools (2014). Directory. *iSchools.org*. Retrieved from <http://ischools.org/members/directory/>.
- Nguyen, T. (2008). Open doors and open minds: What faculty authors can do to ensure open access to their work through their institution. SPARC/Science Commons White Paper, 1-15. Retrieved from https://sparcopen.org/wp-content/uploads/2016/01/opendoors_v1_0.pdf.
- Rucinski, T. L. (2015). The elephant in the room: Toward a definition of grey legal literature. *Law Library Journal*, 107(4), 543-559. Retrieved from <http://www.aallnet.org/mm/Publications/llj/LLJ-Archives/Vol-107/no-4/2015-26.pdf>.
- Schöpfel, J. & Lipinski, T. A. (2012). Legal aspects of grey literature. *The Grey Journal*, 8(3), 137-153.
- Schöpfel, J. & Prost, H. (2013). Degrees of secrecy in an open environment: The case of electronic theses and dissertations. *ESSACHESS – Journal for Communication Studies*, 6(2), 65-86. Retrieved from <http://www.essachess.com/index.php/jcs/article/view/214>.

- Schwartz, M. (2012, Apr. 13). Directory of open access books goes live. *Library Journal*. Retrieved from <http://lj.libraryjournal.com/2012/04/academic-libraries/directory-of-open-access-books-goes-live/#>.
- Suber, P. (2015). Open access overview: Focusing on open access to peer-reviewed research articles and their preprints. *Earlham.edu*. Retrieved from <http://legacy.earlham.edu/~peters/fos/overview.htm>.
- Shieber, S. & Suber, P., eds. (2017). *Good practices for university open-access policies* (wiki). Harvard Open Access Project. Retrieved from bit.ly/goodoa.
- Shieber, S. & Suber, P., eds. (2015). *Good practices for university open-access policies*. Harvard Open Access Project. Retrieved from <https://cyber.harvard.edu/hoap/sites/hoap/images/Bestpracticesguide-2013.pdf>.
- Shieber, S. & Suber, P., eds. (2013). *Good practices for university open-access policies*. Harvard Open Access Project. Retrieved from <https://cyber.harvard.edu/hoap/sites/hoap/images/Bestpracticesguide-2013.pdf>.
- U.S. News and World Report [U.S. News] (2017). Best grad schools rankings: Library and information studies. *USNews.com*. Retrieved from <https://www.usnews.com/best-graduate-schools/top-library-information-science-programs/library-information-science-rankings>.

Bionotes

Professor Tomas A. Lipinski completed his Juris Doctor (J.D.) from Marquette University Law School, Milwaukee, Wisconsin, received the Master of Laws (LL.M.) from The John Marshall Law School, Chicago, Ill., and the Ph.D. from the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Dr. Lipinski has worked in a variety of legal settings including the private, public and non-profit sectors. He is the author of numerous articles and book chapters; his monographs include, *The Library's Legal Answer Book* co-authored with Mary Minow (2003); *The Copyright Law In The Distance Education Classroom* (2005), *The Complete Copyright Liability Handbook For Librarians And Educators* (2006), and *The Librarian's Legal Companion For Licensing Information Resources And Services* (2012). Recent articles and chapters include, *Click Here to Cloud: End User Issues in Cloud Computing Terms of Service Agreements*, in *Challenges Of Information Management Beyond The Cloud: 4th International Symposium On Information Management In A Changing World, Imcw 2013 (Revised Selected Papers.)*, with Kathrine Henderson, *Hate Speech: Legal and Philosophical Aspects*, in *The Handbook Of Intellectual Freedom Concepts* (2014), in 2013 with Andrea Copeland, *Look before you License: The Use of Public Sharing Websites in building Patron Initiated Public Library Repositories*, *Preservation, Digital Technology & Culture* and in 2012, *Law vs. Ethics, Conflict and Contrast in Laws Affecting the Role of Libraries, Schools and other Information Intermediaries*, *Journal Of Information Ethics*. He has been a visiting professor in summers at the University of Pretoria-School of Information Technology (Pretoria, South Africa) and at the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Lipinski was the first named member of the Global Law Faculty, Faculty of Law, University of Leuven, Belgium, in 2006 where he continues to lecture annually at its Centers for Intellectual Property Rights and Interdisciplinary Center for Law and ICT. He is active in copyright education and policy-making, chairing the ACRL Copyright Discussion Group, a member of the ALA OITP Committee on Legislation Copyright Subcommittee, a member of the Copyright and Other Legal Matters Committee of IFLA and serves as an IFLA delegate to the World Intellectual Property Organization's Standing Committee on Copyright and Other Rights. In October of 2014 he returned to the University of Wisconsin to serve as Professor and Dean of its i-School, the School of Information Studies. Email: tlipinsk@uwm.edu



Katie Chamberlain Kritikos is a PhD student at the University of Wisconsin-Milwaukee School of Information Studies. She graduated from the University of Illinois at Urbana-Champaign with a JD (2009) and MLIS (2010) and received her BA, summa cum laude, in English (2006) from the University of Alabama. Katie researches law and information policy with interests in free speech, privacy, information policy, and scholarly communication. She is a member of the Board of Trustees of the Freedom to Read Foundation. Email: kritikos@uwm.edu



How Open Access Policies Affect Access to Grey Literature in University Digital Repositories *A Case Study of iSchools*

GL19 International Conference on Grey Literature
October 23, 2017 ~ Rome, Italy

Dr. Tomas A. Lipinski, Dean and Professor
Katie Chamberlain Kritikos, Ph.D. Student
School of Information Studies, UW-Milwaukee



Introduction to Open Access

- ▶ **Barriers** associated with traditional compilations of journal articles from commercial publishers are absent.
- ▶ **Embargo** period may apply → → → → **dark archive**.
- ▶ OA applies to publications **traditional** (scholarly articles, proceedings, etc.) and **non-traditional** (grey literature).
- ▶ OA allows unfettered **access** to scholarship and promotes open scholarly **communication**.
- ▶ Access to content in university digital repository often managed with **OA policy**.



(c) 2017 TAL & KCK

Research Methodology

- ▶ **RQ:** What are **best practices** for drafting OA policies that balance copyright with unfettered access to and use of grey literature in university digital repositories?
- ▶ **Case study** of OA policies at iSchools.
- ▶ **Sample** of 5 policies:
 - ▶ Tomas A. Lipinski and Katie Chamberlain Kritikos, Legal and Ethical Implications of Licenses Between US Open Access Journal Publishers and Authors, 9th Qualitative and Quantitative Methods in Libraries International Conference (QQML2017), May 23-26 May, 2017, Limerick, Ireland.
 - ▶ Cross-reference 2017 edition of U.S. News and World Reports top 25 US programs with list of iSchools in North America.
 - ▶ <http://ischools.org/regions/north-american-schools/north-american-directory/>.



(c) 2017 TAL & KCK

Case Study & Sample

School	Location	U.S. News 2017 Rank	OA Policy	Repository
University of Illinois at Urbana-Champaign	Champaign-Urbana, IL	1	Yes, 5/14/2016	Illinois Digital Environment for Access to Learning and Scholarship (IDEALS)
University of Washington	Seattle, WA	2	Proposed draft, 6/1/2016	ResearchWorks Archive
University of North Carolina-Chapel Hill	Chapel Hill, NC	3	Yes, 5/13/2016	Carolina Digital Repository
Syracuse University	Syracuse, NY	4	No, info about OA only	Syracuse University Research Facility and Collaborative Environment (SURFACE)
University of Michigan-Ann Arbor	Ann Arbor, MI	5	No, info about OA only	Deep Blue
University of Texas-Austin	Austin, TX	6	Yes, 10/31/2016	Texas ScholarWorks
Rutgers, The State University of New Jersey-New Brunswick	New Brunswick, NJ	7	Yes, 10/29/2014	Scholarly Open Access at Rutgers (SOAR), RUCore
University of Maryland-College Park	College Park, MD	8	No, info about OA only	DRUM
Indiana University-Bloomington	Bloomington, IN	9	Yes, n.d.	IUScholarWorks Repository
University of Pittsburgh	Pittsburgh, PA	10	Proposed draft, 9/13/2013	D-Scholarship@Pitt

(c) 2017 TAL & KCF

Data Collection & Analysis

- ▶ **Green, Gold, Bronze and Black**...and shades in between.
- ▶ **Variables** derived from Harvard Open Access Project's "Good Practices for University Open-Access Policies" (Shieber and Suber, 2017).
 - ▶ **Administrative:** goal/mission, responsibility, application (must or may, opt-out, embargo).
 - ▶ **Rights:** "ownership"/holdership, rights granted, assignable, open licenses (Creative Commons).
 - ▶ **Works:** deposit version, timing (when to deposit), included and excluded.
 - ▶ **Variables added:** student and exclusion (conflicting license).

(c) 2017 TAL & KCF

Findings

- ▶ **Goal/Mission:** disseminate scholarship / perpetual access.
- ▶ **Responsibility:** Provost/VPAA (2) and library (2), IU: library? (powers to remove content).
- ▶ **Policy Application:** mandatory, opt-out allowed except IU; embargo allowed.
- ▶ Non-exclusive **rights** to institution.
- ▶ **Assignable:** (UI, UT and IU?).
- ▶ **Open license:** UI (OA or link to publisher site) and IU ("may consider" Creative Commons).

(c) 2017 TAL & KCF

Findings

7

- ▶ **Works included:** Limited to "scholarly; or "peer-reviewed" articles (except IU); conference proceedings included: UT and Rutgers.
- ▶ **Student works** included: Rutgers and IU, UI via General Rules.
- ▶ **Version:** final, post peer-review except (IU).
- ▶ **Timing of deposit:** not indicated (UI and UNC), no later than publication date (UT and Rutgers), all versions (IU).



(c) 2017 TAL & KCK

Findings

8

- ▶ **Works Excluded:**
 - ▶ Incompatible **license agreement** entered pre OA Policy adoption.
 - ▶ **Classroom** or **pedagogical** materials (UNC, Rutgers).
 - ▶ **Books** (UNC, UT, Rutgers) and Book **Chapters** (UT).
 - ▶ Conference **Posters** (UT).
 - ▶ **General scholarly repository** (UT).
 - ▶ **Dynamic** resources (IU).
 - ▶ **Unlawful** content (IU).
 - ▶ Illinois per General Rules: articles "that fall **outside the scope**."



(c) 2017 TAL & KCK

Who Holds the Copyright?

9

- ▶ **Statutory:** Work Made for Hire, 17 U.S.C. § 101.
 - ▶ 17 U.S.C. § 204(a): signed, written transfer requires identification of the work.
 - ▶ General policy insufficient: *Manning v. Board of Trustees of Community College District No. 505 (Parkland College)*, 109 F.Supp.2d 976 (C.D. Ill. 2000).
- ▶ **Case law:** *Hays v. Sony Corporation of America*, 847 F.2d 412 (7th Cir. 1988).
- ▶ Separate **Copyright/IP Policy:** copyright to scholar.
 - ▶ UI is the only institution that references such policy in its OA document!!!!
- ▶ Authors alerted to **copyright issues** with publication: scholarly communications site (CIC/BTAA or SPARC).



(c) 2017 TAL & KCK

Conclusion

10

- ▶ Except IU there is **bifurcation** of OA and institutional repository policy.
- ▶ Except for UI the OA policies reviewed place **copyright** with the author-scholar, but the derivation of that right is not indicated, nor is information regarding **retention** of copyright during publication included.
- ▶ General inclusion consistent/exclusion less consistent.
 - ▶ Inclusion of **student authors** inconsistent.
 - ▶ Inclusion of peer-reviewed **proceedings** inconsistent.
- ▶ **Embargo** allowed and **waiver** granted.



(c) 2017 TAL & KCK

Recommendations

11

- ▶ Connect OA Policy with broader **institutional repository policies**.
- ▶ OA Policy and **copyright**:
 - ▶ Include **author rights awareness** (CIC/BTAA or SPARC addendum).
 - ▶ Reference **Copyright/IP Policy**.
- ▶ Consider making **scholarly repository policy** mandatory for certain categories of grey literature: working/white papers, reports, proceedings if not in OA repository, and of course ETDs.
- ▶ Consider **Gold OA** as discipline and campus culture allow.
- ▶ Next Step: analysis of university OA repository EULAs.

(c) 2017 TAL & KCK

Thank You!

12

Tomas Lipinski, tlipinsk@uwm.edu
 Katie Chamberlain Kritikos, kritikos@uwm.edu
 School of Information Studies
 University of Wisconsin-Milwaukee
 2025 E. Newport Avenue, NWQB 3rd Floor
 Milwaukee, Wisconsin, USA 53211



(c) 2017 TAL & KCK

Law, Liability, and Grey Literature: Resolving Issues of Law and Compliance

Daniel C. Mack, University of Maryland, United States

Grey literature faces a number of unique challenges when facing issues of liability and compliance with local, national, and international law. Conference and symposium proceedings, white papers, reports, newsletters, and other forms of grey literature often do not have the same legal support as journals, monographs, and other, more formal publications. This can create problems of compliance and liability. Grey publications often include works authored, edited, translated, compiled, or otherwise modified by many people affiliated with multiple institutions, sometimes in several countries. In the process of producing and distributing grey literature, its producers may be confused about, or even unaware of, legal issues such as copyright ownership, authors' and editors' rights, official affiliation, licensing, and distribution. As a result, grey publications may often be at odds with local, national, and international laws and regulations. This in turn may have the unfortunate and unintended consequence of making the producers of grey literature legally liable for damages. This paper proposes a model for identifying potential legal problems, obtaining proper legal counsel, and limiting liability for authors, editors, and distributors of grey literature.

Like other publications, grey literature exists within multiple systems of law and regulation. These systems may include institutional policies and regulations, laws at every level, and the formal and informal networks of researchers, authors, editors, distributors, and others affiliated with grey publications. This paper presents a model based on systems theory to approach the problem. The model offers practical means by which producers of grey literature can identify and address these issues before they become problems, and thus minimize noncompliance and avoid liability.

The resources necessary for implementing this model are usually minimal. The primary expense may be the use of human resources affiliated with sponsoring institutions. This may include administrators, legal counsel, and other personnel not directly involved with the production of grey literature. Other resources, such as Creative Commons licenses, are available at no cost. The model presents a means of identifying and budgeting for costs that are unavoidable.

The result of implementing the model will be publication of grey literature that is in compliance with local, national, and international law, as well as with institutional policies and regulations. This will minimize or eliminate liability for violation of such laws and regulations by researchers.

Bionote

Daniel C. Mack is Associate Dean for Collection Strategies and Services, University Libraries at the University of Maryland in College Park, where he provides leadership in policy creation and implementation, strategic planning, program development, and assessment for library collections. He is also responsible for coordinating copyright and licensing issues for faculty produced publications. His previous positions include Tombros Librarian for Classics and Ancient Mediterranean Studies and Head of the Arts and Humanities Library at Penn State, and Library Director at the Dauphin County (PA) Prison. Mack has advanced degrees in library science and ancient history and has taught college courses in ancient history, Roman archaeology, classical literature and Latin grammar. Recent publications include work as co-editor of the Association of College and Research Libraries' monograph *Interdisciplinarity and Academic Libraries*, as consulting editor for Brill's *New Pauly: Encyclopaedia of the Ancient World* and as author of the "Language, Linguistics and Philology" section of the American Library Association's *Guide to Reference Sources*. Mack's current research interests include interdisciplinarity in the twenty-first century academy, assessment of library collections and services, and Roman civilization in the age of Caesar Augustus. When he has time, Mack plays the viola da gamba and cello. Email: dmack@umd.edu



19th International Conference on Grey Literature
National Research Council of Italy
23-24 October 2017
Rome, Italy

Law, liability, and grey literature: resolving issues of law and compliance

Daniel C. Mack, Associate Dean, Collection Strategies and
Services
University of Maryland Libraries
College Park, Maryland, USA





Compliance challenges to GL

- Local, national, and international law
- Institutional policy
- Standards
- Multiple types and formats of publications, including media
- Multiple types of authors (sole, group, corporate) from a variety of nations and institutions
- Copyright and permissions
- Obtaining official affiliation
- Licensing
- Distribution



Consequences of noncompliance

- Unable to distribute legally
- Author and editor opt-out
- Legal liability, lawsuits, and possible financial consequences
- Mistrust among collaborators

Identifying systems

Regulatory

Local & national law
International law
Institutional policy
Standards

Personnel

Authors
Editors
Distributors
Other

Content

Text
Data
Media
Other



Identifying interactions

Regulatory

Local & national law
International law
Institutional policy
Standards

Personnel

Authors
Editors
Distributors
Other

Content

Text
Data
Media
Other



Human resources for compliance

- Researchers, authors and editors
- Administrators
- Legal counsel
- Print, web and other developers
- There may be a cost associated with use of human resources; calculate time and cost, and include in operating or grant budget

Resources for compliance: copyright and standards

- European Union: <https://ec.europa.eu/digital-single-market/en/policies/copyright>
- Italy: <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:1941-04-22:6331vig=>
- United Kingdom: https://www.copyrightservice.co.uk/copyright/p01_uk_copyright_law
- United States: copyright.gov
- International Standards Organization (ISO): iso.org

Creative Commons:

<https://creativecommons.org>

“Creative Commons provides free, easy-to-use copyright licenses to make a simple and standardized way to give the public permission to share and use your creative work—on conditions of your choice.”

- Choose from a variety of licenses controlling use and commercialization:
<https://creativecommons.org/share-your-work/>

SPARC: <https://sparcopen.org>

"SPARC (the Scholarly Publishing and Academic Resources Coalition) works to enable the open sharing of research outputs and educational materials in order to democratize access to knowledge, accelerate discovery, and increase the return on our investment in research and education."

- Author Addendum to maintain Open Access rights: <https://sparcopen.org/our-work/author-rights/>
- Alternative OA Publishing Models: <https://sparcopen.org/our-work/alternative-publishing-models/>

Creating compliance: identifying system interactions & solutions

Identifying interactions

- What are the regulatory obligations?
- Who are the personnel involved?
- What content is involved?

Identifying solutions

- Copyright, licenses and standards
- Research, administrative, legal, technical support personnel
- Regulatory and technical requirements

Conclusion: results of compliance

Addressing regulatory, personnel, content, and format challenges will result in distribution of grey literature that is:

- In compliance with local, national, and international law
- Meets requirements institutional policies and regulations
- Supports needs of all personnel involved

This will minimize or eliminate liability for researchers, distributors, and others involved in production and distribution of grey literature.

Questions or comments?

Daniel Mack: dmack@umd.edu



THE
NEW YORK
ACADEMY
OF MEDICINE

HEALTHY CITIES.
BETTER LIVES.

URBAN HEALTH UNCOVERED: The Grey Literature Report

**SUBSCRIBE,
SEARCH,
DISCOVER AT:**

GreyLit.org or
greylithelp@nyam.org

Enhance the impact of your research with contemporary, comprehensive urban health resources, selected from more than 750 non-commercial publishers from around the world.

- Researched and curated to give you a unique, inside track on urban health issues.
- Indexed and described to make it easy to discover resources across multiple disciplines.

Uncovering hidden resources for researchers, health professionals, librarians and the public.

NYAM.org

The New York Academy of Medicine

1216 Fifth Avenue | New York, NY 10029

Indexing grey multilingual literature in General Practice in the era of Semantic Web

Marc Jamoulle, Department of General Practice, University of Liège, Belgium

Melissa P. Resnick, University of Texas, Health Science Center at Houston, United States

Ashwin Ittoo, HEC Management School, University of Liège, Belgium

Robert Vander Stichele, Heymans Institute of Pharmacology, University of Ghent, Belgium

Elena Cardillo, Institute of Informatics and Telematics, National Research Council, Italy

Julien Grosjean and **Stefan Darmoni**, D2IM, University of Rouen, France

Marc Vanmeerbeek, Department of General Practice, University of Liège, Belgium

Sharing the results of research with General Practitioners (GPs) is crucial for the survival of the discipline of General Practice / Family Medicine (GP/FM).

The production of abstracts in GP/FM exceeds 15,000 per year worldwide. Each abstract often represents two years of work for its authors and is expressed in local languages. Only 45% of them are published in indexed medical journals.

Usual indexation systems like MeSH are not multilingual nor adapted to the particular field of GP/FM. Consequently, these abstracts are lacking bibliographic control and more than half of the research presented by GPs at congresses is lost.

Considering the absence of appropriate domain-specific terminologies or classification systems, we propose a new multilingual indexing system.

The existing International Classification of Primary Care (ICPC) is currently used for clinical purposes and has now been expanded with a taxonomy related to contextual aspects (called Q-Codes) such as education, research, practice organization, ethics or policy in GP/FM, currently not captured. The set is proposed under the name Core Content Classification in General Practice (3CGP).

The aim is to facilitate indexing of GP/FM specific scientific work and to improve performance in information storage and retrieval for research purposes in this field.

Research Method/Procedure: Using qualitative analysis, a corpus of 1,702 abstracts from six GP/FM-related European congresses was analyzed to identify main themes discussed by GPs (as continuity, accessibility or medical ethics), handled in a domain-specific taxonomy called Q-Codes and translated in 8 languages. In addition, a methodology for building a lightweight ontology (in OWL-2) was applied to Q-Codes, adding object and datatype properties to the hierarchical relations, including mapping to the MeSH thesaurus, Babelnet (www.babelnet.org) and Dbpedia. Finally, the ICPC-2 in 19 languages and Q-Codes in 8 languages have been integrated in a healthcare terminology service (www.hetop.eu/q) with a companion website (<http://3cgp.docpatient.net>)

Anticipated Results of the Research: The creation and the on-line publication of this multilingual terminological resource for indexing abstracts and for facilitating Medline searches could reduce loss of knowledge in the domain. In addition, through better indexing of the grey literature (congress abstracts, master's and doctoral thesis), we hope to enhance the accessibility of research results of GP/FM domain and promote the emergence of networks of researchers. First result of experimental implementations of the new indexing system will be presented.

Indication of costs related to the project: This project has not been funded. 3CGP is placed under Attribution-Non-Commercial-Share-Alike 4.0 International (CC BY-NC-SA 4.0)

Indexing grey multilingual literature in General Practice in the era of Semantic Web

Marc Jamoulle¹, Melissa P. Resnick², Ashwin Ittoo³, Robert Vander Stichele⁴, Elena Cardillo⁵, Julien Grosjean⁵, Stefan Darmoni⁵, and Marc Vanmeerbeek¹

¹ Department of General Practice, University of Liège, Belgium
² School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA
³ HEC Management School, University of Liège, Belgium
⁴ Heymans Institute of Pharmacology, University of Ghent, Belgium
⁵ Institute of Informatics and Telematics, National Research Council, Italy
⁶ Department of Information and Medical Informatics (D2IM), University of Rouen, France

Bibliography ; see text

BACKGROUND

The complex world of reference of family medicine stands between anthropology and bio-sciences

The patient doctor relationships is a heavy consumer and producer of health information

Productions of GPs are a silo of grey literature

There is a need for a specific indexing system, also fit for automatic coding

BACKGROUND

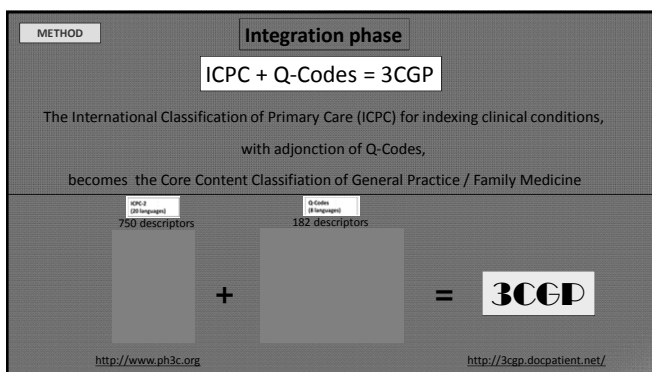
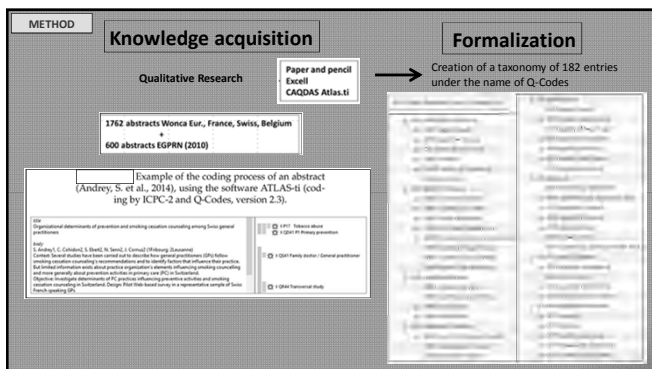
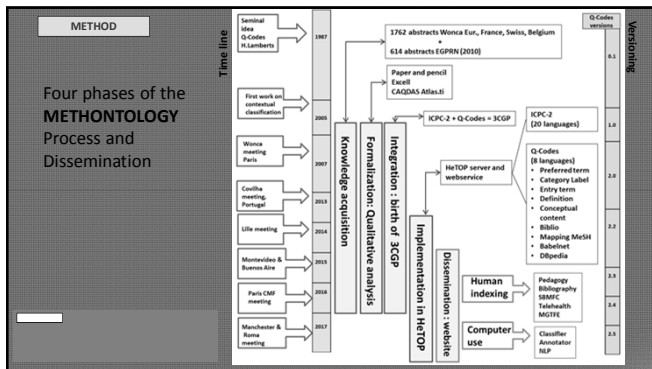
We are presenting the Q-Codes, a taxonomy of contextual issues met by family doctors, complementary to ICPC, an existing Classification.

Q-Codes could be considered as a lightweight ontology ready to be used by humans but also in the semantic web context, and to be exported in Web Ontology Language (OWL)

The multilingual classes of the classification could be individually reached through Unique Resource Identifiers (URIs).

Aim

- To improve annotation of grey literature in primary care,
- To facilitate indexing of congress abstracts and theses
- To improve the searchability of repositories for these information artefacts.



[illegible]

RESULTS SECONDARY

Indexing of Grey Literature

Indexing of master thesis in GP/FM in Belgium and Portugal
Indexing of Brazilian and Uruguayan projects

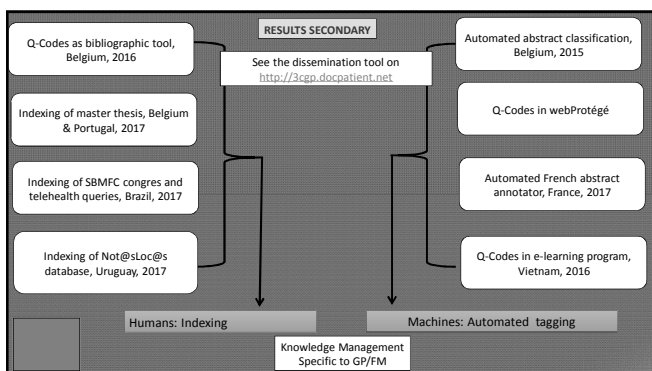
STUDY BY FOCUS GROUP ON THE INVOLVEMENT OF THE FAMILY PHYSICIAN IN OCCUPATIONAL HEALTH PROBLEMS: PERCEPTION OF FAMILY PHYSICIAN AND PATIENTS.

- What would be the objective of the research? Medico-legal problem QD6, Work problem Z05, certificate of illness ICD-2 Process; 62.001 sick leave document
- What context? Family medicine QS41
- What target? Occupational medicine, concept missing in Q-Codes, use QO4
- What method? Focus group is QR6

Coding: QD6; Z05; 62.001; QS41; QO4; QR6

Example of coding : extract of the guide for trainees
Dep Gen Pract. University of Liege. 2016

User guide available on <http://3cgp.docpatient.net>



DISCUSSION	FUTURE RESEARCH
• 3CGP, a new indexing system to be used in GP/F available in OWL	• Q Codes update needed for missed and emergent themes
• Allow teaching of Gp/FM	• Translations (ongoing)
• Experiences of human indexing ongoing	• Reproducibility and Interdoctorvariation
• First step of automatic annotation ongoing	• To populate the light weight ontology
• Choreography by solo dancer	• Integration with upper level ontologies
• Eurocentric	• Automatic annotators in various language
• Good face validity but reproducibility and Inter-doctor variation not tested	• Standardise repositories of abstracts in GP/FM
	• To take place in the distributed data era

ICPC © Wonca

Q-Codes ©m.jamoulle

Slovak Centre of Scientific and Technical Information **SCSTI**

Achieve
your goals
with us



INFORMATION SUPPORT OF SLOVAK SCIENCE

SCIENTIFIC LIBRARY AND INFORMATION SERVICES

- technology and selected areas of natural and economic sciences
- electronic information sources and remote access
- depository library of OECD, EBRD and WIPO

SUPPORT IN MANAGEMENT AND EVALUATION OF SCIENCE

- Central Registry of Publication Activities
- Central Registry of Art Works and Performance
- Central Registry of Theses and Dissertations and Antiplagiarism system
- Central information portal for research, development and innovation - CIP RDI >>>
- Slovak Current Research Information System

SUPPORT OF TECHNOLOGY TRANSFER

- Technology Transfer Centre at SCSTI
- PATLIB centre

POPULARISATION OF SCIENCE AND TECHNOLOGY

- National Centre for Popularisation of Science and Technology in Society

IMPLEMENTATION OF PROJECTS

- National Information System Promoting Research and Development in Slovakia - Access to electronic information resources - NISPEZ
- Infrastructure for Research and Development - the Data Centre for Research and Development - DC VaV
- National Infrastructure for Supporting Technology Transfer in Slovakia - NITT SK
- Fostering Continuous Research and Technology Application - FORT
- Boosting innovation through capacity building and networking of science centres in the SEE region - SEE Science

www.cvtisr.sk

Lamačská cesta 8/A, Bratislava

Preserving and accessing content stored on USB-flash-drives: A TIB workflow

Oleg Nekhayenko

German National Library of Science and Technology, TIB, Germany

Grey Literature has become a corner stone of the supply of scientific and academic literature.

TIB is the German National Library of Science and Technology with a strong focus on Grey Literature. The acquisition of this kind of literature is not a simple process due to the unconventional distribution channels and the cost. For this purpose TIB has its own grey literature acquisition team. In order to keep the data readable and accessible in the long term TIB hosts, administrates and runs a digital preservation system which manages different kind of objects and file formats. Each group of objects needs its own workflow and ingest preparation (pre-ingest). This also applies to grey literature.

Some of the grey literature publications are delivered on USB-flash-drives. In order to catalog them the USB-flash-drive will be accessed by librarians several times. As USB-flash-drives are “write-many” storage devices, accidental write or erase processes could already take place during the access by initially connecting with a PC. This fact leads to the risk of loss or overwriting of rare and expensive data. Furthermore the lifespan of this data storage and the interface dependency are two remaining basic concerns of their long term accessibility in libraries. The minimization of risks of accidental manipulation by direct access to the USB-flash-drives can be achieved through the use of a USB write blocker. This device prevents all write instructions and ensures the safe forensic data transfer. The secure storage of the data from the USB-flash-drive can be done through the separation of data from the storage device. For this purpose TIB has developed a reusable, semi-automated workflow with a strong focus on usability for librarians. Using a script and the write blocker, that prevents the accidental write accesses, the workflow saves the data from USB flash devices via imaging or copying.

This project consists of a market survey (4 person-days), implementation (28 person-days), tests (4 person-days) and launch in team (1 person-days). The price of the write blocker is 500\$.

The main focus of the paper is the detailed description of the project including its technical implementation, discussion of the results and further improvements.

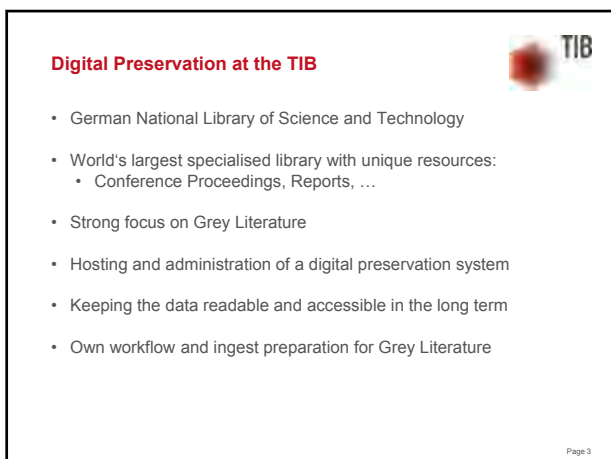
Bionote

Oleg Nekhayenko is research assistant at the German National Library of Science and Technology in the department of digital preservation. He received a bachelor's degree in Information Science and Language Technologies from Heinrich-Heine University Duesseldorf and a master's degree in International Information Management from University of Hildesheim.
Email: Oleg.Nekhayenko@tib.eu









USB-Stick as a storage medium for GL

- Approx. 334 USB-flash-drives with GL in the TIB (2017)
- Increase of the amount in the next years
- „Write-many“ storage device
- Accidental write or erase processes during access
- Risk of loss and overwriting of rare and expensive data
- Not allowed for borrowing and accessing in the reading room
- Not suitable for digital preservation (limited lifespan, manipulation, interface dependency)

Page 4

Content Preservation Strategy**1. Minimization of risks of accidental manipulation during direct access:**

- The use of a USB write blocker
- Blocks the write access to the storage device
- Access only in Read-Only Mode
- Ensures the safe forensic data transfer

**2. Separation of data from the storage device**

- Reusable, semi-automated workflow
- Data saving via imaging or copying

**3. Ingest into digital preservation system**

Page 5

USB write blocker

TABLEAU Forensic USB 3.0 Bridge

**Selection requirements:**

- common model in the IT-Forensic ✓
- Forensic data transfer up to 300MB/Sec (USB 3.0) ✓
- Easy handling ✓
- Also supports USB 1.1/2.0 ✓

Page 6

Separation of data from the data storage



- Batch Script (.bat) for workflow automation
- Guided User Dialog (Instructions and Messages)
- Strong focus on usability for librarians
- Creation of the Image-File and/or copying all the files:
 - Imaging for ensuring the connection between files (if .exe exist)
 - Only File-Copy for independent files (e. g. only pdf)
- Comparison of the MD5-checksums of the image and the original
- LOG-File for error capture (three control mechanisms)

Page 7

Separation of data from the data storage



%EKI-number%/Image/

%EKI-number%/Files/

Page 8

Separation of data from the data storage



```

@echo off
setlocal enabledelayedexpansion

:: Batch Script for separating data from data storage
:: This script will create a directory structure for the data storage
:: and copy the data from the data storage to the specified location.

:: Define the source and target directories
set "SourceDir=%~dp0\SourceDir"
set "TargetDir=%~dp0\TargetDir"

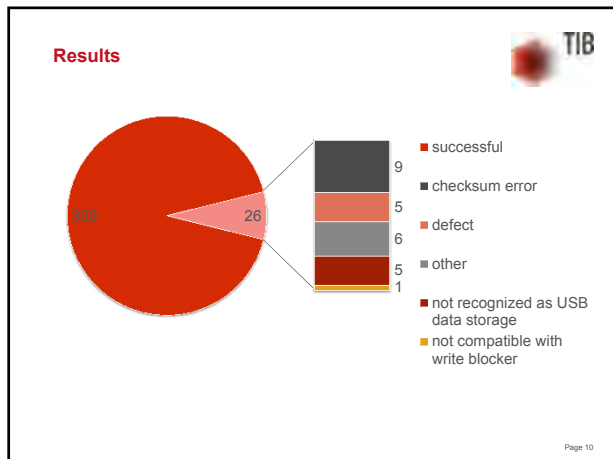
:: Create the target directory structure
mkdir "%TargetDir%\Image"
mkdir "%TargetDir%\Files"

:: Copy the data from the source directory to the target directory
xcopy "%SourceDir%\Image" "%TargetDir%\Image" /E /H /I /Y
xcopy "%SourceDir%\Files" "%TargetDir%\Files" /E /H /I /Y

:: Display the results
echo.
echo =====
echo SourceDir: %SourceDir%
echo TargetDir: %TargetDir%
echo =====
echo.
echo The data has been successfully separated from the data storage.
echo.

```

Page 9



Conclusion and further work

- Data from 92% of USB-Sticks successfully separated
- Different problems occur by 8%
- Data can be transferred by another method from 21 USB-Sticks, errors caused by other reasons
- 4 USB-Sticks lost their data completely (acquisition between 2009-2016), while stored in the library
- 1 USB-Stick was already defect at the time of acquisition
- Failure rate 1,5%
- Digital Preservation of files and images

Page 11

Information Technology Center
University of Vienna

TIB

MORE INFORMATION
www.tib.eu

Contact
Oleg Nekhayenko
T +49 511 762-14518, oleg.nekhayenko@tib.eu

Providing Access to Grey Literature: The CLARIN Infrastructure

Sara Goggi, Gabriella Pardelli, Irene Russo, Roberto Bartolini, and Monica Monachini,
CNR, Istituto di Linguistica Computazionale, "Antonio Zampolli", Italy

"In the electronic age, the World Wide Web has played a major role in making scientific information accessible to a wide audience more rapidly and efficiently. This democratic approach to information dissemination in science is changing the way science is perceived and implemented in our daily lives" (Weintraub, 2000).

Technological process, in particular in the field of computer science, has thus eased access, retrieval and use of information as a consequence of the radical transformation which formats underwent: from papers organized on shelves to electronic files archived on the web. "The Internet has thus had the paradoxical result of making grey literature far easier to access and retrieve than once was the case, but simultaneously making so much available that it is often much harder to find or identify relevant material in the first place" (Hartman, 2006). While in its first days technology, as Hartman states, kept a very quick – and somehow wild – pace in publishing any type of information on line, nowadays there is the need of more sophisticated core technologies and technological building blocks in order to better exploit the huge amount of digital content available on the web. Therefore there is this blossoming of infrastructures, large technological shells which host documentary repositories intended to meet the expectations of a well-educated and demanding audience. The strengthening of these infrastructures at different levels (academic, national, transnational, community, disciplinary, commercial, industrial, etc.) implies a further step in the process of gathering, organizing, managing, preserving and spreading a huge amount of relevant information. "The official definition of 'research infrastructure' refers to structures, resources and services used by a scientific community for carrying out a high-level research in several fields, from the from astronomy, physics, biology, archaeology, to the humanities. At a European level the scientific communities get together in a consortium thus creating infrastructures accessible to all their members and sharing the same resources" (Monachini, Frontini, 2016). Infrastructures stimulate new research avenues, relying on the comparison, re-use and integration into current research of the outcomes of past and on-going field and laboratory activity. Such data are scattered amongst diverse digital collections and datasets, unpublished reports (grey literature), and in publications.

Given this scenario, the authors – who deal with documentation, digitalization and language technologies for the Humanities since years now – focus on an important European research infrastructure called CLARIN (Common Language Resources and Technology Infrastructure) for assessing the traceability of grey literature within it.

On 1st October 2015 Italy became the 16th Full Member of CLARIN ERIC (European Research Infrastructure Consortium). CLARIN-IT is the Italian node of CLARIN, whose grand vision it shares (<http://www.clarin-it.it/en/content/about>).

The *Virtual Language Observatory* VLO [<https://vlo.clarin.eu/search?0>] is the central repository of the CLARIN infrastructure which allows to discover language resources with a facet browser (that is, filtered by semantic categories, such as *Language*, *Collection*, *Resource type*, *Modality*, *Format*, *Keyword*, *Availability*, *Search options*), and advanced query syntax.

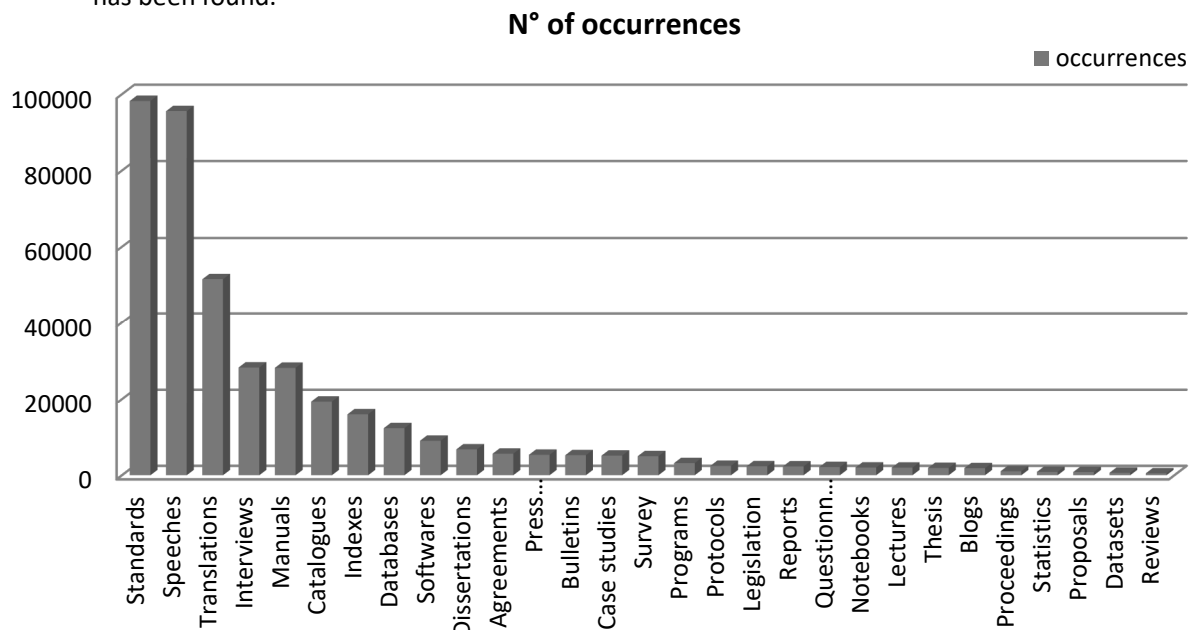
By simply clicking on the term *Language*, the total amount of documents in the various languages is immediately visible from the VLO: the engine has been actually conceived for providing a real time information on the quantity of available resources for that language. All the other categories (*Collection*, *Resource type*, *Modality*, *Format*, *Keyword*, *Availability*, *Search options*) are searchable in cascade.

Grey Literature in CLARIN

The idea is to use the VLO for inquiring about the quantity of grey literature which can actually be found in CLARIN. A mapping between the terminological resources contained in the *GreyNet International 1992-2017* website [www.greynet.org] – and in particular in the

GreySource Index > Document Types in Grey Literature – and those retrievable from the VLO has been performed.

The graph below represents the number of occurrences for each type of document which has been found.



This work will provide a map of the documentation archived in the CLARIN infrastructure, whose purpose is to share language resources¹ produced and managed in the various European countries but finally merged into the CLARIN data centers for allowing access, interoperability, reuse and preservation of scientific documentation as well as Grey Literature.

Essential References

- Rudasill, Lynne M. (2015). Changing dimensions of libraries in the internet era: Grey literature and scholarly communication. In Sanjay Kataria John Paul Anbu K Shri Ram Richard Gartner, *4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services*, Proceedings. Institute of Electrical and Electronics Engineers, Inc. SBN: 978-1-4799-7999-8 (PRINT) . Pages 53-57. doi={10.1109/ETTLIS.2015.7048171}.
- Carroll, B.C. & Cotter, G.A. (1997). A new generation of grey literature: the impact of advanced information technologies, *Pub Res Q* (1997) 13: 5. doi:10.1007/s12109-997-0014-z.
- Weintraub, Irwin,(2000) "The impact of alternative presses on scientific communication", *International Journal on Grey Literature*, Vol. 1 Issue: 2, pp.54-59, doi: 10.1108/14666180010327195
- Banks, M.A. (2006) . Towards a continuum of scholarship: The eventual collapse of the distinction between grey and non-grey literature, *Pub Res Q* (2006) 22: 4. doi:10.1007/s12109-006-0002-8
- Hartman, Karen A.(2006). Social Policy Resources for Social Work: Grey Literature and the Internet, *Behavioral & Social Sciences Librarian*, 25:1. Pages 1-11. doi = {10.1300/J103v25n01_01}.
- Monachini M., Frontini F. (2016). CLARIN, l'infrastruttura europea delle risorse linguistiche per le scienze umane e sociali e il suo network italiano CLARIN-IT. *IJCoL - Italian Journal of Computational Linguistics vol. 2, n. 2, December 2016 "Special Issue on Digital Humanities and Computational Linguistics"*. Pages 11-30. <http://www.aaccademia.it/scheda-libro?aaref=895>
- FLReNet, Fostering Language Resources Network. ECP-2007-LANG-617001. Annex 1.Description of Work (Thematic Network).

¹ "The term 'Language Resources' refers to (usually large) sets of language data and descriptions in machine readable form, to be used in building, improving or evaluating natural language, speech or multimodal algorithms or systems. Typical examples of LRs are written, spoken, multimodal corpora, lexicons, grammars, terminologies, multimodal resources, ontologies, translation memories, but the term may also be extended to include basic software tools for their acquisition, preparation, annotation, collection, management and use. The creation and use of these resources span several related but relatively isolated disciplines, including NLP, information retrieval, machine translation, speech, and multimodality. There is the need today to broaden the definition of LRs, i.e. to re-define the "extension" of the term and recast its definition in the light of recent scientific, methodological-epistemological, technological, social and organisational developments in the application fields of content processing/access/understanding/creation, Human-Machine, Human-Human & Machine-Machine communication, and the corresponding areas from which the theoretical underpinnings of these application fields emerge (linguistics, cognitive science, AI, robotics). The extension of LRs seems to be indispensable to ensure long-lasting credibility. To achieve this, the LRs community must "liaise" with the "new" communities hinted at by the new fields above and draw a new evolving picture of existing/available/future resources following the extended definition" (FLReNet, Fostering Language Resources Network, 2007).

Bionotes

Sara Goggi is a technologist at the Institute of Computational Linguistics "Antonio Zampolli" of the Italian National Research Council (CNR-ILC) in Pisa. She started working at ILC in 1996 working on the EC project LE-PAROLE for creating the Italian reference corpus; afterwards she began dealing with the management of several European projects and nowadays she is involved with organisational and managerial activities mainly concerning international relationships and dissemination as well as organization of events (e.g. LREC conference series). Currently one of her preminent activities is the editorial work for the international ISI Journal Language Resources and Evaluation, being its Assistant Editor. Since many years (from 2004) she also carries on research on terminology and since 2011 - her first publication at GL13 - she is working on topics related with Grey Literature. Email: sara.goggi@ilc.cnr.it



Gabriella Pardelli was born at Pisa, graduated in Arts in 1980 at the Pisa University, submitting a thesis on the History of Science. Since 1984, researcher at the National Research Council, Institute of Computational Linguistics "Antonio Zampolli" ILC, in Pisa. Head of the Library of the ILC Institute since 1990. Her interests and activity range from studies in grey literature and terminology, with particular regard to the Computational Linguistics and its related disciplines, to the creation of documentary resources for digital libraries in the Humanities. She has participated in many national and international projects including the recent projects:- BIBLOS: Historical, Philosophical and Philological Digital Library of the Italian National Research Council, (funded by CNR); - For digital edition of manuscripts of Ferdinand de Saussure (Research Programs of Relevant National Interest, PRIN - funded by the Ministry of Education, University and Research, MIUR). Email: gabriella.pardelli@ilc.cnr.it



Roberto Bartolini - Expertise on design and development of compilers of finite state grammars for functional analysis (macro-textual and syntactic) of Italian texts. Expertise on design and implementation of compilers of finite state grammars for analysis of natural language texts producing not recursive syntactic constituents (chunking) with specialization for Italian and English languages. Skills on acquiring and extracting domain terminology from unstructured text. Skills on semi-automatic acquisition of ontologies from texts to support advanced document management for the dynamic creation of ontologies starting from the linguistic analysis of documents. Email: roberto.bartolini@ilc.cnr.it



Monica Monachini is a Senior Researcher at CNR-ILC. Field of expertise: computational linguistics, computational lexicography, semantics, lexical semantics, language resources, ontologies, lexicon, terminologies, metadata, validation, methods for retrieving information in different areas (biology, environment, civil protection, oceanography, social media, humanities and social sciences, ...), infrastructural issues related to language resources. Active in many standardisation activities for harmonising lexical information. Involved and responsible of the Pisa team in many international projects for language engineering. Over the last years, she has published articles in the field of lexical resources and information extraction in different areas. Currently, she focused her activities on digital humanities. Member of various Scientific Committees; UNI delegate for ISO/TC37/SC4. Email: Monica.Monachini@ilc.cnr.it



Data Visualization of a Grey Literature Community: A Cooperative Project

Roberto Bartolini, Sara Goggi, Gabriella Pardelli, and Irene Russo, ILC-CNR Italy
Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands

In 2012, GreyNet published a page on its website and made accessible the first edition of IDGL, International Directory of Organizations in Grey Literature¹. The latest update of this PDF publication was in August 2016, providing a list of some 280 organizations in 40 countries worldwide that have contact with the Grey Literature Network Service. The listing appears by country followed by the names of the organizations in alphabetical order, which are then linked to a URL.

This year GreyNet International marks its Twenty Fifth Anniversary and seeks to more fully showcase organizations, whose involvement in grey literature is in one or more ways linked to GreyNet.org. Examples of which include: members, partners, conference hosts, sponsors, authors, service providers, committee members, associate editors, etc.

This revised and updated edition of IDGL will benefit from the use of visualization software mapping the cities in which GreyNet's contacts are located. Behind each point of contact are a number of fields that can be grouped and cross-tabulated for further data analysis. Such fields include the source, name of organization, acronym, affiliate's job title, sector of information, subject/discipline, city, state, country, ISO code, continent, and URL. Eight of the twelve fields require input, while the other four fields do not.

The population of the study was derived by extracting records from GreyNet's in-house, administrative file. Only recipients on GreyNet's Distribution List as of February 2017 were included. The records were then further filtered and only those that allowed for completion of the required fields remained. This set of records was then converted to Excel format, duplications were removed, and further normalization of field entries took place. In fine, 510 records form the corpus of this study. In the coming months, an in-depth analysis of the data will be carried out - the results of which will be recorded and made visually accessible.

The expected outcome of the project will not only produce a revised, expanded, and updated publication of IDGL, but will also provide a visual overview of GreyNet as an international organization serving diverse communities with shared interests in grey literature. It will be a demonstration of GreyNet's commitment to research, publication, open access, education, and public awareness in this field of library and information science. Finally, this study will serve to pinpoint geographic and subject based areas currently within as well as outside of GreyNet's catchment.

1. <http://www.greynet.org/internationaldirectory.html>

Bionotes

Roberto Bartolini, see page 62

Sara Goggi, see page 62

Gabriella Pardelli, see page 62

Dominic Farace, see page 25

Jerry Frantzen, see page 26

Strategies for Teaching and Learning about Grey Literature, including the Dissemination and Exchange of Information

Marcus Vaska, Alberta Health Services, Canada

Dean Giustini, University of British Columbia Library, Canada

Hana Vyčítalová, National Library of Technology, Czech Republic

Introduction Creating greater awareness of grey literature has been a key goal of GreyNet since its founding. At the 18th International Conference on Grey Literature (GL18), held at the New York Academy of Medicine in 2016, Lynne Rudasill, Librarian and Associate Professor at the University of Illinois at Urbana-Champaign, presented her research on teaching and learning about grey literature for a specific group, namely, library school students in the United States. Rudasill's findings showed that many LIS students do not receive much instruction about grey literature but are introduced to this material while on the job. To address this issue, a cross-curricular approach was adopted, as suggested in 2011 by Debbie Rabina, associate professor at Pratt Institute School of Library and Information Science, to more "closely reflect the broad scope of grey literature, [which should] be more prominently placed in the curriculum of library and information programs" (Rudasill, 2017). Further, at GL18, participants were encouraged to provide their views on four questions: 1) how can the grey literature community address questions such as how one can improve the understanding of grey literature through formal learning? 2) how should learning objectives be established? 3) how can the topic be expanded to more general audiences and 4) what methods can be used to disseminate information on grey literature more broadly?

Research Questions/Aims The mandate of the GreyNet LIS Education and Training Committee reflects the various projects underway in the grey literature community to enhance education and training. Using Rudasill's work as a guide, the Information Delivery subgroup will investigate methods of engagement and information exchange in grey literature by focusing on the following three activities: 1) compile a list of educational and promotional activities, along with practical experiences in the area of grey literature across various disciplines, 2) conduct an environmental scan by means of a literature search to uncover prior work on learning objectives and teaching methods around grey literature; and 3) explore how grey literature is represented in Google Scholar (a cross-comparison with Google).

Discussion/Conclusions As a means of recording our thoughts and analysis with regards to the three activities of our group, we created a communication space using Blogger software. Our poster will pull themes from the GreyNet Information Delivery blog, <http://greynetinformationdelivery.blogspot.ca/>, to develop a set of definitions when discussing grey issues, topics and strategies around research in grey literature. Our goal is to create greater awareness of the importance of including grey literature alongside white literature (published commercial articles retrieved from major databases). Developing a set of strategies for teaching and learning about grey literature, including finding better ways to disseminate and exchange this information, is an iterative and long term goal for our group.

Bionotes

Marcus Vaska is a librarian with the Knowledge Resource Service, Alberta Health Services, responsible for providing research and information support to staff at an Alberta Cancer Care Centre. A firm believer in embedded librarianship, Marcus engages himself in numerous activities, including instruction, patient engagement, and research consultation with numerous teams at this facility. An advocate of the Open Access Movement, Marcus' current interests focus on strategies for creating greater awareness of grey literature via various information dissemination and exchange pursuits. Email: mmvaska@ucalgary.ca

Hana Vyčítalová studied Information science and librarianship at the Charles University in Prague. Since 2012 she works in the National Library of Technology in Prague (Czech Republic) in Digital National Library of Technology Department. Currently she is partnership network manager of the National Repository of Grey Literature. She is interested in grey literature, open access, research data, enhanced publications and free licences. She is coordinator of the Conference on Grey Literature and Repositories in Czech Republic. Email: hana.vycitalova@techlib.cz

Strategies for Teaching and Learning About Grey Literature, Including the Dissemination and Exchange of Information

Marcus Vaske¹, Dean Giustolisi², Hana Vycitalová³
¹Tom Baker Cancer Centre, Canada; ²Norcross General Hospital, Canada; ³National Library of Technology, Czech Republic
marcus.vaske@utoronto.ca, dean.giustolisi@utoronto.ca, hana.vycitalova@nlk.cz

INTRODUCTION/BACKGROUND

At GL18, participants provided their views on four questions pertaining to grey literature (Rudasill, 2016):

- How can we, in the GL community, address how to improve global awareness and understanding of GL?
- Can formal learning objectives be established? What are the Stages of Inquiry?
- Can they be expanded to general audiences, and
- What methods can we use to disseminate information on grey literature more broadly?

Stages of Inquiry

Our blog:
<http://greyetinformationdelivery.blogspot.ca>

Selected Educational/Promotional Activities - Canada
[Grey Notes: Current Awareness Blog in Current Awareness International - Grey Literature](#)
[Leading Edge II: Searching for Grey Literature](#)
[Role of Grey Literature in Academic Library Collections](#)

Selected Educational/Promotional Activities - Czech Republic
[Conference on Grey Literature and Technologies](#)
[National Repository of Grey Literature](#)

RESEARCH QUESTIONS/AIMS

To investigate methods of engagement and information exchange by focusing on three activities:

- I. Listing educational and promotional activities, along with practical experiences in the realm of grey literature across various disciplines.
- II. Doing an environmental scan and literature search to uncover prior work on learning objectives and teaching methods around grey literature, and
- III. Exploring how grey literature is represented in Google Scholar (a cross-comparison with Google).

DISCUSSION & FUTURE DIRECTIONS

- In the digital era, much grey literature can be found on the Web by using search engines. Librarians the world over can raise awareness about how to locate GL more systematically.
- As part of our teaching "How to search for GL", librarians can emphasize definitions of GL, challenges in GL searching and the coverage of search engines beyond Google.
- We plan to share our findings about this topic again in 2018.

GL18 COVERAGE BY GOOGLE & GOOGLE SCHOLAR

- GL is reported to have in excess of 144 million citations, pointing to millions of cited grey literature documents unavailable in controlled databases.
- Numerous studies have estimated the number, subject coverage and characteristics of documents available in GL.
- We plan to perform cross-comparison test searches to estimate the number of GL documents indexed by GL, while providing updates about its coverage, crawling ("indexing") practices & interface features.
- Search topics will be representative of scholarly disciplines in government, academic and business sectors.

How we learn, teach, and describe Grey Literature

Background: The Rudasill Study 2016

- How to improve *global awareness* of GL
- Formal learning objectives & **Stages of Inquiry**?
- Expanded to general audiences, and
- Methods to disseminate information on GL?

Committee Research Aims

- I. **LIST** educational, promotional, practical experiences
- II. **COMPLETE** environmental scan and literature search
- III. **EXPLORE** how GL literature is covered in Google Scholar (a cross-comparison with Google).

The International GreyNet Community 2017

91 authors from 21 countries

- Algeria
- Australia
- Austria
- Canada
- Czech Republic
- France
- Germany
- Greece
- India
- Italy
- Japan
- Netherlands
- Norway
- Pakistan
- Russia
- Slovakia
- South Korea
- Sweden
- United Kingdom
- USA
- Zimbabwe

How We Teach, Learn, & Describe Grey Literature in 46 Words



Selected Educational Promotional Activities

◦ CANADA

[Grey Horizon Current Awareness Blog in Cancer Care](#)

[HLWIKI International – Grey Literature](#)

[Leading Edge II: Searching for Grey Literature](#)

[Role of Grey Literature in Academic Library Collections](#)

◦ CZECH REPUBLIC

[Conference on Grey Literature and Repositories](#)

[National Repository of Grey Literature](#)

GL in Google & Google Scholar

- **GOOGLE SCHOLAR (GS)** has 144 million+ citations and **grey literature** documents **not in** controlled databases.
- **MANY STUDIES** estimate the type of documents in GS.
- **WE WILL COMPARE** GL documents indexed by GS and GOOGLE, coverage, crawling (“indexing”) practices & interface features.
- Searches re: government, academic & business sectors.

Future Directions

- Today, *grey literature* is found everywhere on the Web
- Librarians are raising awareness about how to locate it
- In teaching “how to search for GL”, librarians can:
 - emphasize definitions of GL
 - challenges in searching for it, and
 - coverage of search engines beyond Google.
- We will share our findings again in 2018

References

- Barger, V., & Labrecque, L. (2013). An integrated marketing communications perspective on social media metrics. *Int J Integrat Marketing Communications*, 64-76.
- Giustini, D., & Boulos, M. N. K. (2013). Google Scholar is not enough to be used alone for systematic reviews. *Online journal of public health informatics*, 5(2), 214.
- Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature. *Journal of Informetrics*, 11(3), 823-834.
- Lin, Y., & Vaska, M. (2013). Creating and assessing a subject-based blog for current awareness within a cancer care environment.. *The Grey Journal*, 9(1), 7-13.
- Rudasill, L. (2017). Teaching and learning about grey literature. *GL 18 Conference Posters 18(1)*, 114-115

National Repository of Grey Literature (NRGL)

NU
SL

NRGL is
digital
repository
for grey
literature

Free
online
access

Features

Provider:

National Library of Technology
Prague, Czech Republic

Records:

over 400,000 records

Collection provenance:

Czech Republic

Partners:

over 130 organizations (Academy of Science,
Public Research Institutions, Universities, State
Offices, Libraries, NGOs etc.)

International Cooperation:

OpenGrey, OpenAire, ROAR, OpenDOAR, BASE

Goals

- Central access to grey literature and the results of research and development in the Czech Republic
- Support of science, research and education
- Systematic collection of metadata and digital documents
- Long-term archiving and preservation
- Cooperation with foreign repositories

What else?

Conference on Grey Literature
and Repositories

<http://nrgl.techlib.cz/conference/>

Informative Web pages

<http://nrgl.techlib.cz>

www.nusl.cz

NTK
Národní technická knihovna
National Library of Technology

**NU
SL**
národní
úložiště
šedé
literatury

Collecting Grey Literature – Institutional Repository versus National Aggregator

Petra Černohlávková and Hana Vyčítalová,
NTK, National Library of Technology, Czech Republic

The Czech National Library of Technology (NTK) provides two digital repositories – the National Repository of Grey Literature (NRGL) and the NTK Institutional Digital Repository (IDR). NRGL's primary is providing access to grey literature as well as long-term archiving and preservation of grey literature from various institutions in the Czech Republic. The IDR was created for collecting and archiving of employee-generated content and other documents, including grey literature, connected with the library and its services. Our poster highlights the differences between collecting grey literature at the institutional level (through the institutional repository) and at the national level. What commonalities and differences do they have? What problems do they solve? Differences include not only overall conceptions and document types, but also methods for collecting, legal issues and standards as well as functionality and options. Thanks to our experiences in managing both types of repositories, we define general differences, obstacles, and development possibilities. Information presented here, including a mode for cooperating at the institutional or national level, is useful for all institutions planning to start collecting (not only) grey literature at the institutional or the national level even at cooperating institutional model/level.

Bionote

Petra Černohlávková studied Information science and librarianship at the Charles University in Prague. She is currently working in the National Library of Technology (NTK) in Prague in Digital National Library of Technology Department. She is a content coordinator of the National Repository of Grey Literature and of the Institutional Repository of NTK. Email: petra.cernohlavkova@techlib.cz
Hana Vyčítalová studied Information science and librarianship at the Charles University in Prague. Since 2012 she works in the National Library of Technology in Prague (Czech Republic) in Digital National Library of Technology Department. Currently she is partnership network manager of the National Repository of Grey Literature. She is interested in grey literature, open access, research data, enhanced publications and free licences. She is coordinator of the Conference on Grey Literature and Repositories in Czech Republic. Email: hana.vycitalova@techlib.cz

Grey Literature LibGuides or LibGuides about Grey Literature: A 2-Continent Environmental Scan of Common Themes & Trends

Kathleen Noma Carlson, University of Arizona College of Medicine, USA
Joachim Schöpfel, Université de Lille, Sciences Humaines et Sociales, France
Marcus Vaska, Alberta Health Services, Canada

Introduction - The past decade has seen a continuous and exponential rise in LibGuides, a content management and library knowledge-sharing system well-represented in academic institutions worldwide. As a type of grey literature (officially receiving this distinction from the Gernot community in 2016), LibGuides certainly exhibit many grey literature traits, including easy online access, providing up-to-date information, promoting self-learning, and containing information that is often overlooked and neglected. According to the LibGuides Community website, <http://community.libguides.com/>, 542,190 LibGuides presently exist worldwide, authored by 137,825 librarians from 4,907 institutions in 70 countries.

Aims/Objectives - The mandate of the GreyNet LIS Education and Training Committee reflects on the various informational pursuits being used in the grey literature community in order to enhance education and training in the field of grey literature. Using the LibGuides Community website as an information portal, an environmental scan was conducted across Canada, the United States, and Europe in the spring of 2017 to locate standalone grey literature LibGuides or subject-specific LibGuides devoting considerable attention to grey material. LibGuides selected for the environmental scan were chosen geographically (to ensure balance across provinces, states, and countries), and were evaluated and critically appraised according to the AACODS checklist, https://dspace.flinders.edu.au/jspui/bitstream/2328/3326/4/AACODS_Checklist.pdf

Discussion/Conclusion - Our poster will thus describe the evidence gathering process undertaken to determine appropriate LibGuides used for analysis, extrapolating common themes and trends universal across all LibGuides. These include a definition of grey literature, information on how to search for grey literature, grey literature document types, lists of resources, along with any evidence of social media inclusion. Creating a checklist of necessary elements that should be present across all LibGuides will create great awareness and spread the word for any academic institution that may wish to follow suit and create either a separate or embedded LibGuides devoted to grey literature for their subject area.

Bionotes

Kathleen Noma Carlson, MLS, AHIP, is the associate and education librarian at the University of Arizona College of Medicine – Phoenix. She has worked here since 2012, and her position is to support the College of Medicine medical students, pathway scholars, faculty and staff. Carlson has spent nearly 20 years as a librarian. She received her Master's degree in library science from Texas Women's University. Prior to coming to the UA College of Medicine – Phoenix, she worked as a hospital librarian for 12 years at John C Lincoln Hospitals (HonorHealth) and at Arizona State University. As the health sciences librarian, she has taught nursing and allied health students.
 Email: kncarlso@email.arizona.edu



Joachim Schöpfel is senior lecturer at the Department of Information and Library Sciences at the Charles de Gaulle University of Lille 3 and Researcher at the GERiiCO laboratory. He is interested in scientific information, academic publishing, open access, grey literature and eScience. He is a member of GreyNet and euroCRIS. He is also the Director of the National Digitization Centre for PhD Theses (ANRT) in Lille, France.

Email: joachim.schopfel@univ-lille3.fr

Twitter: <https://twitter.com/schopfel>

Marcus Vaska see page 64



Grey Literature LibGuides or LibGuides about Grey Literature: A Two Continent Environmental Scan of Common Themes & Trends

Kathleen Norma Carlson, University of Arizona College of Medicine, Phoenix, Arizona, USA; Joachim Schöpfel, Université de Lille Sciences Humaines et Sociales, Lille, France; Marissa Vaska, Alberta Health Services, Calgary, Alberta, Canada
joachim.schopf@univ-lille.fr marissa.vaska@ahs.ca kathleen.carlson@arizona.edu

Arguments for Using Grey Literature *

1. Grey literature is an important source of information due to the comprehensiveness of the content that grey literature provides. Because commercial publishers are looking to make a profit on the materials they publish, they often overlook niche research areas that serve smaller populations. Grey literature is one way to search for information in emerging or less popular research areas.
2. Grey literature can sometimes be more current than commercially published information. It does not go through the potentially time-consuming peer-review process undertaken by commercial publishers, and therefore has a quicker turnaround time for dissemination.
3. The importance of grey literature can vary between disciplines. While the health and medical fields tend to favor traditional publication methods through well-known scholarly publishers, other fields like disaster management tend to prefer the grey literature route.

Methodology

Using the LibGuides Community website as an information portal, an environmental scan was conducted across Canada, the United States, and Europe in the spring of 2017 to locate standalone grey literature LibGuides or subject-specific LibGuides devoting considerable attention to grey material. LibGuides selected for the environmental scan were chosen geographically (to ensure balance across provinces, states, and countries), and were evaluated and critically appraised according to the AACODS checklist.

Discussion/Conclusion

Our poster will thus describe the evidence gathering process undertaken to determine appropriate LibGuides used for analysis, extrapolating common themes and trends universal across all LibGuides.

These include a definition of grey literature, information on how to search for grey literature, grey literature document types, lists of resources, along with any evidence of social media inclusion.

Best Practices Checklist:

- Grey literature definition
- Document types
- Searching for grey literature
- List of resources
- Value/Importance of grey literature
- How to evaluate/appraise grey literature

Importance of LibGuides for Grey Literature

LibGuides well-represented in multiple disciplines

LibGuides received designation from GreyNet as a recognized type of grey literature in 2016

Currently, there are 542,190 LibGuides authored by 137,825 librarians from 4,907 institutions in 70 countries

Methodology

- LibGuides Community website as an information portal
- Environmental scan across Canada, the United States, and Europe
- Standalone grey literature LibGuides or subject-specific LibGuides devoting considerable attention to grey material
- Evaluation according to the AACODS checklist
- Several LibGuides, selection of 12 for poster

Sheridan College, Canada: A Sample Best-Practice LibGuide



Best Practices Checklist

- Grey literature definition
- Document types
- Searching for grey literature
- List of resources
- Value/importance of grey literature
- How to evaluate/appraise grey literature

Development of Visualization Search Service Model using Keyword Social Network for R&D Report

YongJu Shin, Ki Seok Choi, ChulSu Lim, Jiseong Son, and Kwang-Nam Choi, NTIS Center, KISTI, Hwan Gue Cho, Dept. of Computer Science and Engineering, PUSAN National University, Korea

As the amount of research data produced and used has increased rapidly due to advances in technology, there have been various attempts to handle it with Big Data and to provide useful results to researchers. Among these, large-scale documents such as R&D reports need a more efficient way to identify the contents and obtain relevant document information in a short time.

In order to quickly understand the contents of a document, visualizing the content of a R&D report in a single book will help to understand whether it is a desired R&D report. The model that visualizes and provides the contents of the R&D report proposed in this study is a document summary model that reflects the spatio-temporal characteristics and shows the dynamic change of words developed according to the content of the document. By applying this to the R&D report, the researcher can select a particular section in the document's time base to identify it in the form of a keyword graph.

In conclusion, this model can quickly identify the content of R&D reports that researchers can't understand at one time. Combining this model with the search service of existing research reports, it will help you quickly find the documents you need and the contents of R&D reports.

Acknowledgments

This research was supported by Maximize the Value of National Science and Technology by Strengthen Sharing/Collaboration of National R&D Information funded by the Korea Institute of Science and Technology Information (KISTI).

Bionotes

Ki Seok Choi is a principal researcher of NTIS Center, Korea Institute of Science and Technology Information(KISTI), Korea. He received the B.S. degree from the school of Computer Science and Statistics at Seoul National University, in 1988, and , M.S. degree from the school of Information and Communication Engineering at Korea Advanced Institute of Science and Technology, in 1997, PH.D. degree from the school of Computer Engineering at Chungnam National University 2013, respectively. His research interests include information retrieval, database and distributed data interoperability. Email: choi@kisti.re.kr

Kwang-Nam Choi received his M.S degree in computer engineering from Chungnam National University, Daejeon, Korea, in 1994. Since 1994 he has been with KISTI, where he is a principal member of the research staff. Email: cjchae@kisti.re.kr



Korea Institute of Science and Technology Information (KISTI)

English version - <http://en.kisti.re.kr/>

* Vision

World-class information research institute creating values for customers

* Main functions



* Management and service of Korean R&D reports

KISTI exclusively manages, preserves, and serves Korean R&D reports for citizens and government officials. It provides Korean R&D reports and their information with National science & Technology Information Service (NTIS) and National Discovery for Science Leaders (NDSL).

*Contact information

KISTI email address: hcpark@kisti.re.kr

Headquarters: Tel : +82-42-869-1004, 1234 Fax: +82-42-869-0969

Classification Assistance for Korean R&D Project Using Machine Learning Techniques

Jiseong Son, Kwang-Nam Choi, and Chul-Su Lim, NTIS Center, KISTI;

Yu-Chul Jung, Dept. of Computer Engineering, Kumoh National Institute of Technology, Korea

We devise a classification assistance for Korean R&D project using machine learning techniques. It enables recommendation of suitable class code for classification of research area with R&D project based on Korean national science and technology standard classification hierarchy. The classification of R&D project is important for management and distribution of information, efficient management for human resource, measurement of statistical data and efficient planning projects in research area. Currently, class codes for classifications should be manually input by researcher in Korean R&D project information registration system. However, an accurate classification of R&D project has a few issues. First, it has many kinds of codes in Korean national science and technology standard classification system that is consist of 3 levels. Second, a convergence research area doesn't have a well-defined boundary. Therefore, researchers spend a lot of time for finding classification code that is suited for their research area and cause some errors such as wrong input code. In addition, it needs an expert of each research area to modify wrong input code and requires more time and cost for analyzing R&D project information. In order to resolve these issues, we implement classification assistance using SVM algorithm and experiment with machine learning techniques using R&D project information that consist of project title(Korean, English), keywords(Korean, English), research goal, abstract, expected effect, classification code, and so on. For evaluation of classification, we measure an accuracy. Using the result of measurement, we improve accuracy and usability of classification for R&D area. Furthermore, it can reduce time for verification of input codes.

Acknowledgments

This research was supported by Maximize the Value of National Science and Technology by Strengthen Sharing/Collaboration of National R&D Information funded by the Korea Institute of Science and Technology Information (KISTI).

Bionote

Kwang-Nam Choi received his M.S degree in computer engineering from Chungnam National University, Daejeon, Korea, in 1994. Since 1994 he has been with KISTI, where he is a principal member of the research staff. Email: cjchae@kisti.re.kr

International Nuclear Information System INIS

*organizing the world's information
on nuclear science and technology
and making it universally accessible
for peaceful uses*

over 150 Member States and
international organizations

millions of citations and
abstracts published worldwide

hundreds of thousands of full text
non-conventional 'grey' literature

multilingual thesaurus in Arabic,
Chinese, English, French, German,
Japanese, Russian, Spanish



www.iaea.org/inis

IAEA
International Atomic Energy Agency

GreyGuide Portal and Repositories “Sharing knowledge as early as possible”

Stefania Biagioni and Carlo Carlesi, ISTI-CNR

Alessia Barsotti and Andrea Pardini, UNSC-ISTI-CNR, Italy

The GreyGuide - Repository and Guide to Good Practices and Resources in Grey Literature was launched in 2013 as a collaborative effort between GreyNet International and ISTI-CNR. It aims at * Offering information professionals, practitioners, and students common ground for good practices and resources in grey literature; * Providing support for all stakeholders working on Grey Literature related research issues; * Being consistent with recommendations made in the Pisa Declaration on Policy Development for Grey Literature Resources as well as Open Science guiding principles.

.... The European Commission sees the shift to an open science system as a source of competitive advantage for Europe. Open science covers the cycle of research from conceptualizing to analyzing and publishing ... It embraces all types of scientific knowledge, from research data to journal articles to presentation slides, and all manner of stakeholders: researchers, funders, policymakers, citizens, enterprises, and publishers.

Open science is a new approach to the scientific process based on cooperative work, coupled to new tools for collaboration, and new routes for knowledge diffusion through online digital technologies. Open science entails a shift from the standard practice of publishing research results in scientific journals, towards sharing all available data and knowledge at the earliest stages of the research process. It requires a move from ‘publishing as fast as possible’ to ‘sharing knowledge as early as possible’ ...

This year activities dealing with the GreyGuide have focused on developing a new version of its Portal and designing a thematic repository – a work still in progress. The focus is geared to ‘Open Access Knowledge’, which includes aspects of Open Access, Open Data, and Open Source. In brief, it adheres to an Open Science Vision.

Two new open access sections have been added to the Portal in an effort to make available diverse types of documents, namely: Slides Presentations, Conference Proceedings, Program books, GreyNet In-house Publications, Info Adverts; and Training materials.

1. *Next-generation metrics: Responsible metrics and evaluation for open science.* Report of the European Commission Expert Group on Altmetrics. Directorate-General for Research and Innovation. 2017

Bionotes

Stefania Biagioni see page 18

Carlo Carlesi, graduated in Computer Science, worked since 1970 at the IEI (now ISTI) of the CNR in Pisa. He is currently a Research Associate of the Institute ISTI and he is involved in the following projects: PUMA - Publication Management. The Digital Library service allows public access (when permitted) through Internet to the published documents produced by CNR Organizations. And GreyGuide, portal and repository of good practice and resources in the field of grey literature. Email: carlo.carlesi@isti.cnr.it



Alessia Barsotti graduated in Digital Humanities at the University of Pisa. She is currently involved as a member of the library staff in the Project of the National Civil Service (UNSC) at ISTI-CNR in Pisa, Italy. Her research interests include Digital Library Management Systems, scholarly communication services, published and Grey Literature cataloguing, as well as web publishing/e-publishing. Alessia is nearing completion of her master's degree in Electronic Publishing at the University of Pisa. Email: alessia.barsotti@isti.cnr.it



Andrea Pardini graduated as accountant, commercial expert and software developer at the Istituto Tecnico Commerciale, ITC Pacinotti in Pisa, Italy. Currently, he carries out activities at the ISTI-CNR as a member of the National Civil Service Project (UNSC). He is a technical collaborator of the Library and is interested in New Technologies and Digital Library Management Systems. His expertise focuses on the realization of web portals and services. In addition, Andrea is interested in the study of weather forecast, managing a professional weather station and processing weather forecasts for Italy, www.retemeteoamatori.it Email: andrea.pardini@isti.cnr.it



Library, Information Science & Technology Abstracts™ with Full Text

Available via EBSCOhost

The definitive professional information
resource designed for librarians and
information specialists...

*Library, Information Science & Technology Abstracts™
with Full Text* is an indispensable tool for librarians
looking to stay current in this rapidly evolving field.

Comprehensive content includes:

- Full text for more than 270 journals and nearly 20 monographs
- Indexing for more than 550 core journals, 50 priority journals and nearly 125 selective journals
- Includes books, research reports, proceedings and author profiles
- Access to 6,800 terms from reference thesauri
- Coverage extends back as far as the mid-1960s

Subject coverage includes:

- Bibliometrics
- Cataloging
- Classification
- Information Management
- Librarianship
- Online Information Retrieval
- And much more...

Contact EBSCO Publishing to learn more about
Library, Information Science & Technology Abstracts™ with Full Text,
or to request a free trial.

Phone: 800.653.2726

Email: request@ebscohost.com

www.ebscohost.com

A Facet-based Open and Extensible Resource Model for Research Data Infrastructures

Leonardo Candela, Luca Frosini, and Pasquale Pagano,
Istituto di Scienza e Tecnologie dell'Informazione (ISTI) "Alessandro Faedo"
National Research Council (CNR), Italy

Background: Research Data play a key role in our society. They include both “primary dataset”, i.e. data genuinely produced, as well as “derived datasets”, i.e. datasets resulting by processing existing datasets. Their management requires dedicated e-Infrastructures and a description of the entire set of “resources” surrounding each dataset (e.g. other datasets, services the dataset has been produced with or suitable for “consuming” the dataset, entities responsible for the dataset). To realise this, it is key to implement an information system capable to represent datasets as well as the rest of resources associated with it. The “descriptions” attached to such “entities” should not be prescribed a priori, rather they should be open and extensible thus to enable diverse actors (being them publishers or consumers) to annotate each entity with specific features.

Objective: In this work it is proposed a Resource Model capable of representing the different resources involved in the “management” of a research dataset, i.e. the dataset itself as well as the rest of services and entities worth being considered to properly deal with the dataset. The relationships linking the described datasets and associated resources are captured by giving the possibility to describe the specific linking purpose thus to capture the role each associated resource has in the specific context of the dataset. The Resource Model is specifically conceived to deal with the very evolving scenarios arising in the context of Research Data Infrastructures.

Methods: The proposed Resource Model has been tested on a real scenario to prove that it is capable of capturing all aspect to be tracked in such a context.

Results: The study leads to the definition of a basic Resource Model based on two different entity concepts: *Resource* and *Facet*. The former is any entity worth managing in the context, the latter is used to capture a portion of the specification of a Resource. Moreover, two different types of basic relations are needed: the first between Resource and Resource called *isRelatedTo*, and the second between a Resource and a Facet called *consistsOf*. By relying on these very basic constructs it is possible to serve the needs arising in real scenarios.

Conclusion: To capture the information needs arising in Research Data Infrastructure when managing Research Data it is needed an open and extensible model. This work proposed one based on few basic concepts, and described how it has been exploited to serve the needs arising in specific contexts.

Bionotes

Leonardo Candela is a researcher at Networked Multimedia Information Systems (NeMIS) Laboratory of the Italian National Research Council - Institute of Information Science and Technologies (CNR-ISTI). He graduated in Computer Science in 2001 at University of Pisa and completed a PhD in Information Engineering in 2006 at University of Pisa. He was involved in various EU-funded projects and was a member of the DELOS Reference Model Technical Committee and of the OAI-ORE Liaison Group. His research interests include Data Infrastructures, Virtual Research Environments, Data Publication, Open Science, Digital Library [Management] Systems and Architectures, Digital Libraries Models, Distributed Information Retrieval, and Grid and Cloud Computing. Email: leonardo.candela@isti.cnr.it

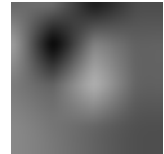


Bionotes CONTINUED

Luca Frosini is Researcher at Networked Multimedia Information Systems (NeMIS) Laboratory of the Italian National Research Council - Institute of Information Science and Technologies (CNR-ISTI). He graduated in Computer Science in 2006 at University of Pisa. Luca was involved in various EU-funded projects (e.g. DILIGENT, D4Science, BlueBRIDGE, PARTHENOS, SoBigData) His research interests include Data Infrastructures, Virtual Research Environments, Software Repositories, Accounting systems, Distributed Information System and Grid and Cloud Computing. He is one of the top most contributor of gCube software, the open-source platform for the management and operation of scientific data infrastructures. Email: luca.frosini@isti.cnr.it



Pasquale Pagano is Senior Researcher at the Networked Multimedia Information Systems Laboratory of the "Istituto di Scienza e Tecnologie della Informazione A. Faedo" (ISTI) of the Italian National Research Council (CNR). He received his M.Sc. in Information Systems Technologies from the Department of Computer Science of the University of Pisa (1998), and the Ph.D. degree in Information Engineering from the Department of Information Engineering: Electronics, Information Theory, Telecommunications of the same university (2006). "The aim of my research is the study and experimentation of models, methodologies and techniques for the design and development of distributed virtual research environments (VREs) which require the handling of heterogeneous computational and storage resources, provided by Grid and Cloud based e-Infrastructures, for the management of heterogeneous data sources. I have a strong background on distributed architectures. I participated to the design of the most relevant distributed systems and e-Infrastructure enabling middleware developed by ISTI - CNR. I am currently the Technical Director of D4Science, the Hybrid Data Infrastructure serving scientists in 37 countries, and chief manager of gCube software, the open-source platform for the management and operation of scientific data infrastructures. I am collaborating with the Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources (iMarine). I am serving the BlueBRIDGE European Project as Technical Director, the Parthenos European Project as Service Operation Manager, and the SoBigData Research Infrastructure as Infrastructure Manager." Email: pasquale.pagano@isti.cnr.it

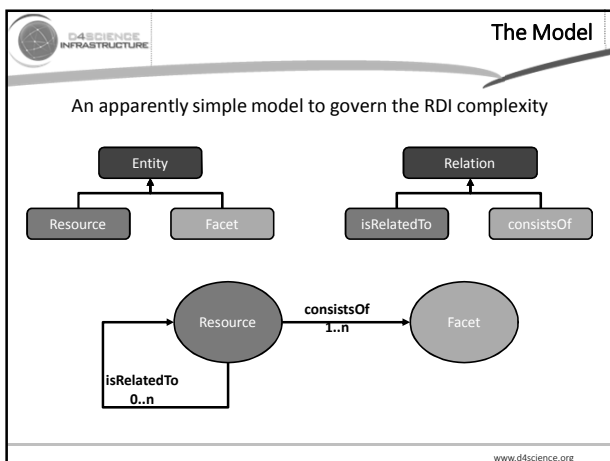


The RDI landscape

Research Data Infrastructure (RDI)
harnesses the accumulating data and knowledge produced by the communities of research, optimizing the data movement across scientific disciplines, enabling large increases in multi- and inter-disciplinary science while reducing duplication of effort and resources, and integrating research data with published literature

- An RDI is intrinsically a system of systems federating data, tools, services, and actors belonging to heterogeneous scientific communities
- An RDI is disassembled into its component systems, each organization and managing their own resources
- An RDI is naturally geographically distributed
- An RDI evolves accordingly to the evolution of the component systems

The IMarine Case www.d4science.org 0



Facet

Facets contribute to "build" a description of a **Resource**

- Every **Facet** is characterized by a number of properties
- Every **Facet**, once attached to a **Resource** profile, captures a specific characterization of the **Resource**

www.d4science.org



Resource

Resources represent
a description of *"thing"* to be managed

- Every **Resource** is characterized and described by a number of **Facets**
- Every **Resource** has no properties

www.d4science.org



Relations

Relations model
a relationship between two entities.

Two typologies of relations are envisaged:

- **isRelatedTo**: establish a relationship among two **Resources**
- **consistsOf**: connect a **Resource** with a **Facet** characterizing it

www.d4science.org



Relations (cont.)

- Any **Relation** has a direction, i.e. a **"source"** (out bound of the relation) and a **"target"** (in bound of the relation).
 - Any relation can be also navigated in the opposite direction
- It is not permitted to define a **Relation** having a **Facet** as **"source"**. In other words:
 - It is not permitted to define a **Relation** connecting a **Facet** with another one
 - It is not permitted to define a **Relation** connecting a **Facet** with a **Resource** (as target)
- A **Facet** instance can be linked (by **consistsOf** or any specialization of it) from different **Resources**
- Every **Relation** is characterized by a number of properties

www.d4science.org

Entities and Relations Specializations

Starting from the simple and powerful model we defined a number of specializations to support the Research Data Infrastructure scenario

- We defined the type of resources
- We defined the relationships among those resources
- We defined the description of those resources through the definition of a number of facets

To manage the openness and the evolution of the RDI, the model can be easily specialized by defining

- new typologies of relations
- new typologies of entities (both resources and facets)

www.d4science.org

RDI : Identified Resources

A Research Data Infrastructure

- manages Services, Software, and Datasets
- operated and provided by Actors
- provisioned by different Sites
- and configured to satisfy the needs of the Scientific Communities

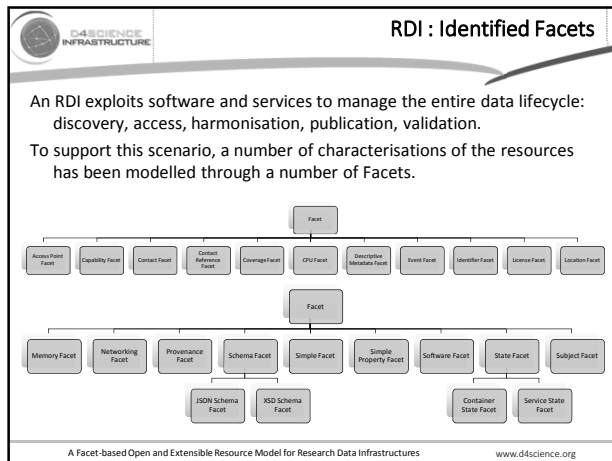
A Facet-based Open and Extensible Resource Model for Research Data Infrastructures

www.d4science.org

RDI : Identified Relations

Relations among resources allow to define the semantic and the behaviour expressed by the relationship

www.d4science.org



D4Science

D4Science is a Hybrid Data Infrastructure

- connecting **+3900 scientists in 44 countries**
- Integrating data from **+50 heterogeneous providers**
- executing **+42,000 models & algorithms/month**
- provides access to over a **billion quality records** in repositories worldwide
- operating with **99,7% service availability**

D4Science hosts **+95 Virtual Research Environments** to serve the biological, ecological, environmental, and statistical communities world-wide.

www.d4science.org

CONTACTS

luca.frosini@isti.cnr.it
leonardo.candela@isti.cnr.it
pasquale.pagano@isti.cnr.it

Technical References
[https://wiki.gcube-system.org/gcube/Facet Based Resource Model](https://wiki.gcube-system.org/gcube/Facet-Based+Resource+Model)

www.d4science.org

D4Humanities: Deposit of Dissertation Data in Social Sciences & Humanities – A Project in Digital Humanities

Joachim Schöpfel, GERiICO Laboratory, University of Lille
Hélène Prost, CNRS, associate GERiICO Laboratory, France

Problem/goal: Following our work on research data and electronic theses and dissertations since 2013, we conduct a new research project in 2017 and 2018 with three objectives – develop the research data management and stewardship on our campus, gain better insight in the nature of research data in social sciences and humanities and produce empirical evidence on the development of dissertations.

Research method/procedure: We will present the preliminary results from our research project D4Humanities.

Anticipated results: The project D4Humanities is part of the Digital Humanities - how to enable the exploration of research data in social sciences and humanities (textual or oral corpus, raw data, images ...) with digital technologies (text and data mining, mapping, visualization ...) to build a new meaning? It is a continuation of the recent research of the GERiICO laboratory and its partners at the University of Lille Humanities and Social Sciences (academic library, SSH graduate school, digitization centre ANRT...) with the objective of accelerating the research data management project in particular for PhD students and young researchers, and of fostering the preparation of an international research project. In particular, the project contains three components:

1. Qualitative survey on behaviours and knowledge in the field of research data with 50 scientists from the University of Lille Social Sciences and Humanities, with a special focus on the FAIR guiding principles of scientific data management and stewardship.
2. The creation of a workflow for the submission of research data related to PhD dissertations (deposit, preservation and dissemination of data via the Nakala service TGIR Huma-Num)
3. Two conceptual studies on the definition and typology of research data in SSH and on the development of dissertations in the environment of eScience and Open Science (content, format, structure, requirements).

The project is conducted with the ISN Oldenburg and other international partners; it will facilitate the creation of a consortium and the elaboration of a larger research project in Digital humanities on the PhD dissertations of the future, with European funding (H2020) or French-German funding (ANR / DFG).

Practical implications: Model for other universities; better understanding of research data in SSH and of the development of dissertations.

Related costs: The project receives funding (€25,000) from the European Institute of Social Sciences and Humanities (MESHS Lille) and from the Regional Government (Conseil Régional Hauts-de-France).

Bionotes

Joachim Schöpfel is senior lecturer at the Department of Information and Library Sciences at the Charles de Gaulle University of Lille 3 and Researcher at the GERiICO laboratory. He is interested in scientific information, academic publishing, open access, grey literature and eScience. He is a member of GreyNet and euroCRIS. He is also the Director of the National Digitization Centre for PhD Theses (ANRT) in Lille, France. Email: joachim.schopfel@univ-lille3.fr

Hélène Prost is information professional at the Institute of Scientific and Technical Information (CNRS) and associate member of the GERiICO research laboratory (University of Lille 3). She is interested in empirical library and information sciences and statistical data analysis. She participates in research projects on evaluation of collections, document delivery, usage analysis, grey literature and open access, and she is author of several publications.

Email: helene.prost@inist.fr



D4Humanities
Deposit of dissertation data in SSH
A project in digital humanities

Joachim Schöpfel
Hélène Prost

 Université de Lille
SCIENTIFICO HUMAINES
ET SOCIÉTÉS

 gérifco

 MESHS
Métiers de l'Édition et de la Science

 Région
Hauts-de-France

Focus on ETDs → Former research

- 2005 Production and findability
- 2013-2015 Accessibility in open access context
- 2014-2016 Data as complementary material
- 2017 *D4Humanities* – workflow and more

GL19 – October 23-24, 2017 – National Research Council of Italy - Roma 2

Survey on research data → Results

- Interest & motivation
 - RDM isn't priority, 27 volunteers only
- Funding agencies
 - 14 participations in research project, but lack of knowledge about guidelines or recommendations for RDM
- Privacy
 - 13 usage of personal data, 6 research protocol submitted to the university's ethics committee
- Standards
 - 8 participants reported assigning codes to their data, 9 drafted a DMP
- Dissemination and sharing
 - 16 agree to share their data with others, 10 deposited data to online server, 2 others intend to do so, only one refuses for security reasons

GL19 – October 23-24, 2017 – National Research Council of Italy - Roma 3

Survey on research data

→ Needs

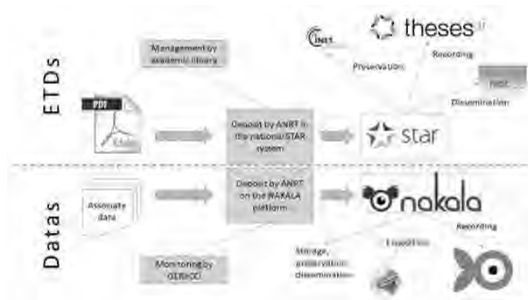
- Advices
 - Querying databases, formatting and naming data
 - Licencing and legal issues, sensitive data
 - Advice from IT department and ethics committee
- Data services
 - Data storage: what data to store, under which formats, on which server, with which guarantees of duration and security
 - Encourage exchanges between researchers and information professionals

GL19 – October 23-24, 2017 – National Research Council of Italy - Roma

4

Data workflow

→ principles



5

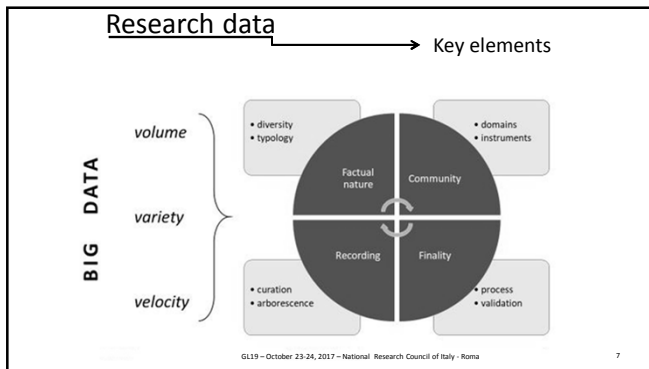
Data workflow

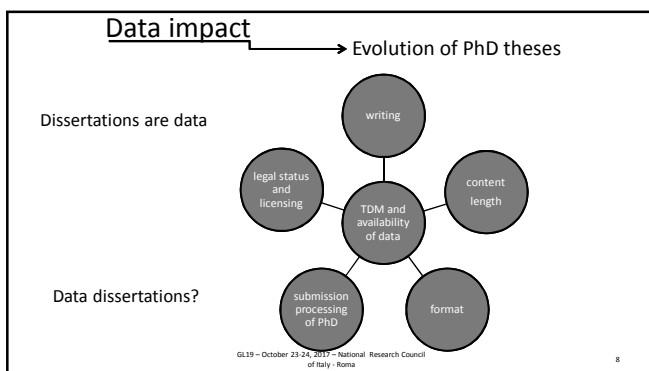
→ Questions

- Granularity
 - A level which makes sense for understanding (validation) and reuse, and to allow deposit of dataset collections with a hierarchical structure
- Data structure and description
 - The Metadata Encoding & Transmission Standard of the Library of Congress
- Identifier
 - The handle system (applied by the Huma-Num infrastructure)
- Legal aspects
 - Students have to sign a disclaimer that they have the rights to upload the datasets on NAKALA
- Quality
 - No criteria of scientific quality, but the question is open (validation by PhD jury)

GL19 – October 23-24, 2017 – National Research Council of Italy - Roma

6





Perspectives

- An overview on ongoing research
- A setting up of a scientific consortium around a core project team
- The preparation of an international research project on new forms of PhD dissertations, with European (H2020) or French-German funding (ANR/DFG)

GL19 – October 23-24, 2017 – National Research Council of Italy - Roma 9



If you are interested:

**TELL US MORE ABOUT YOU,
AND JOIN THE CONSORTIUM**

Contact
joachim.schopf@univ-lille3.fr

Your 7 steps to sustainable data



1. Prepare your data

Select the relevant data files. Check them for privacy aspects and file format against the guidelines issued by DANS.



2. Go to EASY

Log in at <https://easy.dans.knaw.nl>. If you are new to EASY, you will have to register for an account first.



3. Start the deposit procedure

Go to 'deposit your data', select your discipline and click 'start deposit'.



4. Documentation and access level

Describe the dataset and indicate whether it is open access or whether access restrictions apply.



5. Upload your data files

Select your data files and click 'upload dataset'.



6. Submit your data files

Accept the licence agreement and send your dataset to DANS by clicking the 'submit' button.



7. Publication by DANS

DANS will verify the dataset and publish the description you made. Your data have now been sustainably archived and will be accessible to others on a permanent basis under the conditions you specified.



Newsletter:
info@dans.knaw.nl



Twitter:
[@greynetnl](https://twitter.com/greynetnl)



YouTube:
[DANSdataarchiving](https://www.youtube.com/DANSdataarchiving)



Email:
info@dans.knaw.nl



Website:
dans.knaw.nl

Assessing the FAIRness of Data Sets in Trustworthy Data Repositories

Peter Doorn, Data Archiving and Networked Services, DANS-KNAW, Netherlands

Research funding in recent years often comes with the condition to make the resulting data openly available. Just opening up research data is not enough: the data should also be of sufficient quality. In this presentation, I will propose criteria and methods to assess the quality of data sets. Part of the quality assurance can be guaranteed by digital repositories that archive and provide access to data. I will argue that the certification criteria of digital archives and the FAIR data principles for data sets provide a good basis for guaranteeing the quality or “fitness for use” of research data sets.

The core certification offered by the Data Seal of Approval (DSA) and World Data System (WDS) for data repositories, in combination with the FAIR data principles get as close as possible to giving quality criteria for research data. They do not do this by trying to make value judgements about the content of datasets, but rather by qualifying the fitness for data reuse in an impartial and measurable way. By bringing the ideas of the DSA/WDS and FAIR together, we will be able to offer an operationalization that can be implemented in any certified Trustworthy Digital Repository.

In 2014 the FAIR Guiding Principles (Findable, Accessible, Interoperable and Reusable) were formulated. The well-chosen FAIR acronym is highly attractive: it is one of these ideas that almost automatically get stuck in your mind once you have heard it. In a relatively short term, the FAIR data principles have been adopted by many stakeholder groups, including research funders.

The FAIR principles are remarkably similar to the underlying principles of the DSA, which date back to 2005: these specify that the data can be found on the Internet, are accessible (having clear rights and licenses), are in a usable format, are reliable, and are identified in a unique and persistent way so that they can be referred to. Essentially, the DSA presents quality criteria for digital repositories, whereas the FAIR principles target individual datasets.

Bionote

Dr. Peter Doorn is director of DANS. He was co-founder of the Netherlands Historical Data Archive in 1989 and has been active in the domain of digital research data ever since. He is chair of the Science Europe Working Group on Research Data, national representative and vice-chair of the CEESDA General Assembly, former national representative of DARIAH ERIC, former chair of Research Data Netherlands, board member of the Research Data Alliance Organisational Advisory Board, and (board) member of various other national and international data-related organizations. He is editor of the recently founded *Research Data Journal for the Humanities and Social Sciences*.

Email: peter.doorn@dans.knaw.nl



FAIR Data Principles and Grey Literature

Peter Doorn (DANS)

19th International Conference on Grey Literature

Public Awareness and Access to Grey Literature
Rome, October 23-24, 2017



www.dans.knaw.nl
DANS is an institute of KNAW and NWO

What is FAIR data and why is it important?

- Open Science has become a policy priority: not only publications need to be openly accessible, but also other research outputs, especially research data
- Principle: **open if possible, protected if necessary**
- Legitimate restrictions to openness:
 - to protect privacy
 - data creator must be able to publish first
- Openness itself is not enough, data must also be:

Findable	F	
Accessible	A	
Interoperable	I	
Reusable	R	

Do the FAIR principles also apply to Grey Literature?

Everybody loves FAIR!



Everybody wants to be FAIR... But what does that mean? How to put the principles into practice?

FAIR Metrics Group

GOALS:

- Develop a broadly useable framework and tool for FAIR assessment
- Harmonize current ongoing efforts

Expected outcomes:

- Checklist of clearly defined metrics
- Indication of how these metrics can be readily implemented (e.g. self-reporting, automated)
- Proposal(s) for undertaking a FAIR assessment of a digital resource
- One or more implementations for performing an assessment

FAIR Metrics Form

Metric Identifier
Metric Name
To which principle does it apply?
What is being measured?
Why should we measure it?
What must be provided?
How do we measure it?
What is a valid result?
For which digital resource(s) is this relevant?
Examples of their application across types of digital resource
Comment

Can we do it?

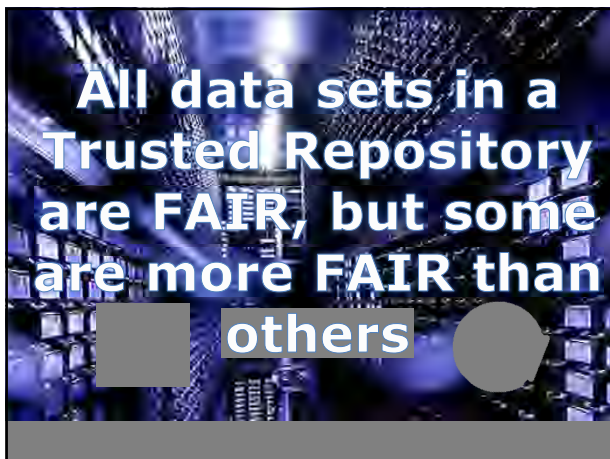
Resemblance Data Seal of Approval – FAIR principles

DSA Principles (for data repositories)	FAIR Principles (for data sets)
data can be found on the internet	Findable
data are accessible	Accessible
data are in a usable format	Interoperable
data are reliable	Reusable
data can be referred to	(citable)

The resemblance is not perfect:

- usable format (DSA) is an aspect of interoperability (FAIR)
- FAIR explicitly addresses machine readability
- etc.

A certified TDR already offers a baseline data quality level



FAIR badge scheme



- Proxy for data "quality" or "fitness for (re-)use"
- Prevent interactions among dimensions to ease scoring
- Consider Reusability as the resultant of the other three:
 - the average FAIRness as an indicator of data quality
 - $(F+A+I)/3=R$
- Manual and automatic scoring

Data Archiving and Metadata Services
DANS

Findable (defined by metadata (PID included) and documentation)

1. No PID nor metadata/documentation
2. PID without or with insufficient metadata
3. Sufficient/limited metadata without PID
4. PID with sufficient metadata
5. Extensive metadata and rich additional documentation available

DANS FAIR
Core



Accessible (defined by presence of user license)

1. Metadata nor data are accessible
2. Metadata are accessible but data is not accessible (no clear terms of reuse in license)
3. User restrictions apply (i.e. privacy, commercial interests, embargo period)
4. Public access (after registration)
5. Open access unrestricted

Interoperable (defined by data format)

1. Proprietary (privately owned), non-open format data
2. Proprietary format, accepted by Certified Trustworthy Data Repository
3. Non-proprietary, open format = 'preferred format'
4. As well as in the preferred format, data is standardised using a standard vocabulary format (for the research field to which the data pertain)
5. Data additionally linked to other data to provide context

Data Archiving and Metadata Services
DANS

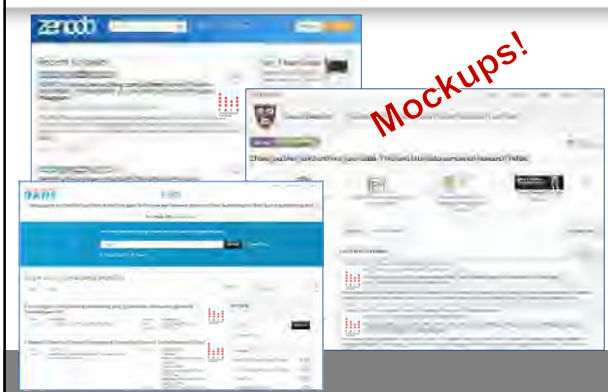
Creating the FAIR Data Assessment Tool



Using an online questionnaire system

Prototype:
<https://www.surveymonkey.com/r/fairdat>

Display FAIR badges in any repository (Zenodo, Dataverse, Mendeley Data, figshare, B2SAFE, ...)



Testing the prototype

- How to assess multi-file datasets?
 - e.g. DANS archives: 35,000 data sets, 4 million files (on average > 100 files per dataset)
- Do the same principles apply to all data types?
 - eg. images, PDFs?
- Operationalizing Reusability?
- Max score for Accessibility of sensitive data?
- Multiple reviewer variance?
- Scores are additive
- Do the metrics reflect the FAIR principles well?
 - Replace DANS metrics by those of Go FAIR metrics group?

Data Archiving and Repository Service
DANS

Distribution of top 20 of file types/formats in DANS EASY archive

How to interpret FAIRness of:

- images
- pdf
- gis-files
- ms-word documents

Does that make sense at all?

Are these also "research data"?

How to assess multi-file datasets?

What does it mean for Grey Literature to be FAIR?

Rank	Extension	File type	N of files	% of total
1	jpg	image/jpeg	2.633.879	64,5%
2	pdf	application/pdf	164.630	4,0%
3	tif	image/tiff	137.408	3,4%
4	csv	data/plain	129.762	3,2%
5	txt	text/plain	122.136	3,0%
6	mid	gis/mid	57.415	1,4%
7	mif	gis/mif	57.029	1,4%
8	cmdi	text/xml	53.054	1,3%
9	tab	text/plain	46.431	1,1%
10	map	gis/map	43.297	1,1%
11	id	application/octet-stream	42.759	1,0%
12	dat	data/	41.671	1,0%
13	dbf	data/database	41.358	1,0%
14	xml	text/xml	35.681	0,9%
15	gz	application/gzip	30.327	0,7%
16	doc	text/ms-word	30.281	0,7%
17	shx	???/octet-stream	21.947	0,5%
18	shp	gis/shape-file	21.946	0,5%
19	gif	image/gif	21.687	0,5%
20	dxf	image/x-dxf	21.262	0,5%
Subtotal			3.753.960	91,9%
> 1200 other formats/extensions			331.708	8,1%
All			4.085.668	100,0%

Data Archiving and Repository Service
DANS

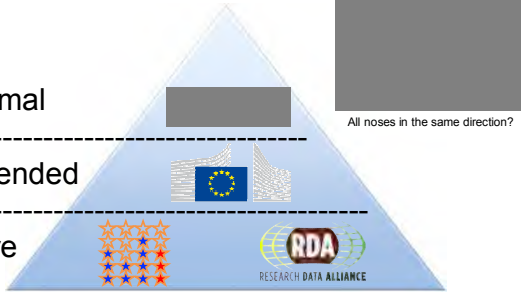
Towards a FAIR Framework?

Analogous to Certification Framework?

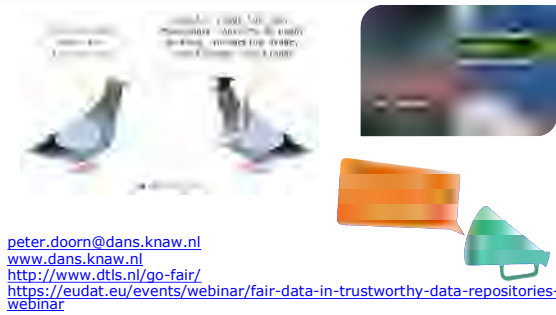
Formal

Extended

Core



Thank you for listening!



Thanks to Ingrid Dillo and Emily Thomas for their contributions



Assessing Data Management Needs and Practices to Enable Research Data Support Services

Plato L. Smith and Jean Bossart, University of Florida Libraries, United States

Researchers are faced with many data management challenges resulting from the need to comply with data management and sharing requirements of funding agencies. While major funding agencies such as the American Heart Association (AHA), National Institutes of Health (NIH), National Science Foundation (NSF), and United States Department of Agriculture (USDA) have data management and sharing requirements guidelines and/or statements, researchers seeking funding struggle with complying, fulfilling, and negotiating data lifecycle management. Data lifecycle management requirements include a.) The key components of a data management plan (DCC, 2013), b.) Key data lifecycle processes (USGS, 2013), c.) Data management lifecycle (UNSW, 2017), and d.) Research data management services (Jones, Pryor, & Whyte, 2013) for scalability and sustainability. The coordination, collaboration, and connection of diverse stakeholders across multiple departments, organizations, and units in support of good data management requires consistent organization, technology, and resources to maintain stability (University of Glasgow Humanities Advanced Technology & Information Institute [HATII], University of London Computer Centre, DCC, & Joint Information Systems Committee [JISC], 2013). The University of Florida (UF) is in pursuit of good data management within and across multiple disciplines.

“It will be difficult to improve your institutional infrastructure without an overall understanding of the data you currently hold and how researchers at your institution are managing their data.” – CARDIO v.2 (Collaborative Assessment of Research Data Infrastructure and Objectives)

A UF data survey (IRB# 201602303) titled “Investigating Data Assets, Management, and Planning at UF” administered from January 2017 through April 2017 yielded results in support of education, support, and training to better assist researchers in meeting the data management and sharing requirements of funding agencies. This presentation will introduce preliminary data survey results that contributed to the development of a NSF Computer and Information Science and Engineering (CISE) Research Initiation Initiative (CRII). The purpose of the CRII is to develop socio-technical (people, policies, technologies, and communities) components to build capacity, infrastructure, and support for expanding and facilitating the culture of good data management across UF communities of practice.

Bionotes

Plato Smith is the Data Management Librarian at the University of Florida with experience in academic research libraries, digital libraries, and data management. He received his doctorate in the field of Information Science from the School of Information within the College of Communication and Information at Florida State University, Florida’s iSchool, Summer 2014. From 2005 to 2012, he was Department Head for the FSU Libraries’ Digital Library where he developed, populated, and managed digital collections in the FSU Libraries’ digital content management system, DigiNole Repository, and electronic theses and dissertations (ETDs) institutional repository. Email: plato.smith@ufl.edu



Jean Bossart is an Associate Engineering Librarian at the Marston Science Library at the University of Florida. She earned a Bachelor’s degree in chemical engineering and a Master’s degree in environmental engineering and worked as a research and development engineer and engineering consultant before joining the library. She is a licensed professional engineer in Florida. Her research interests include engineering education and leadership, and women and underrepresented groups in engineering.

Email: jean.bossart@ufl.edu





**ASSESSING DATA MANAGEMENT NEEDS AND PRACTICES TO
ENABLE RESEARCH DATA SUPPORT SERVICES**

Plato Smith, Ph.D., Data Management Librarian
Jean Bossart, Engineering Librarian
Nineteenth International Conference on Grey Literature
National Research Council of Italy
October 24, 2017

UF George A. Smathers
Libraries
UNIVERSITY of FLORIDA

TABLE OF CONTENTS

1. Background & Context
2. Research Purpose
3. Research Questions
4. Research Design & Methodology
5. Findings
6. Future Plans

BACKGROUND & CONTEXT

Four core resource sets:

1. Workshops on data management planning previously conducted at UF, specifically: **UF Marston Science Library** (9/22/16; 7/11/17), **UF Informatics Institute** (10/24/16; 1/23/17; 1/30/17; 4/28/17), **UF Division of Student Affairs** (6/8/17), **UF Biomedical Sciences** (7/7/17), and **UF Graduate Linguistics Society Seminar** (9/7/17)
2. Data Collected from the UF survey (**IRB# 201602303**) titled "Investigating Data Assets, Management, and Planning at UF" (1/3/17 – 4/31/17)
3. Data Management Use Case – genomics datasets in zenodo (*internal UF Libraries Strategic Opportunity Program grant*)
4. Data Management Instructional Design Internship, fall 2017 – spring 2018 (data management plan learning exercise (**DMPLE**))

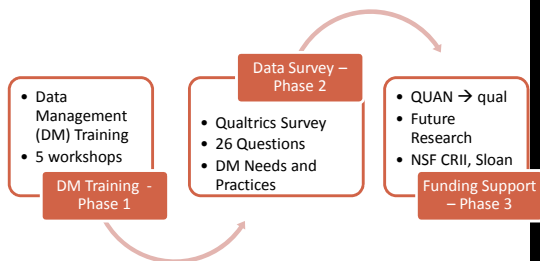
RESEARCH PURPOSE

The purpose of this research project is to investigate researchers current data management practices across multiple disciplines and explore opportunities for improving data management activities where applicable, beneficial, feasible, and reasonable.

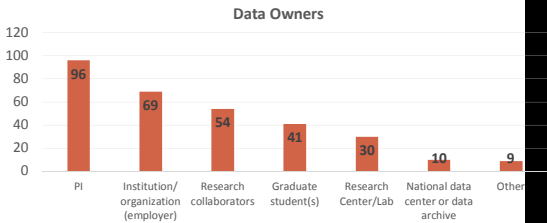
RESEARCH QUESTIONS

1. Who are the owners of data?
2. Who are the managers of data?
3. What are some data management support and training needs?

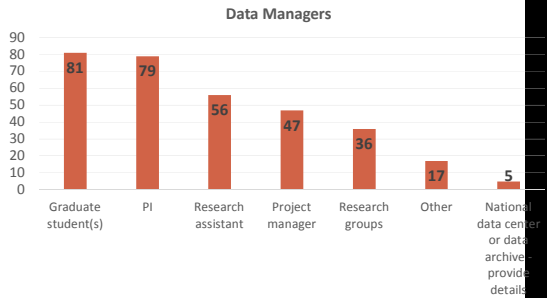
RESEARCH DESIGN & METHODOLOGY



FINDINGS – Q18. WHICH OF THE FOLLOWING ARE OWNERS OF THE DATA?[SELECT ALL THAT APPLY]



FINDINGS – Q17. WHICH OF THE FOLLOWING ARE INVOLVED IN THE MANAGEMENT AND STORAGE OF DATA? [PLEASE SELECT ALL THAT APPLY]



FINDINGS – Q23. WHAT RESOURCES, SUPPORT, AND TRAINING OUTSIDE OF YOUR DEPARTMENT DO YOU FIND USEFUL TO SUPPORT FUNDING AGENCIES DATA MANAGEMENT PLANS REQUIREMENTS? [PLEASE SELECT ALL THE APPLY]

Resources, support, and training outside of your department	Number of responses
Short-term and long-term data storage capacity	66
Training on data management planning, sharing, and preservation	61
Data/digital management system for organizing data	55
Computing expertise or software development	49
Computing capacity for data analysis and data visualization	48
Research data management services (e.g. in house, campus, outsource)	43
Other external expertise (e.g. developer, programmer, statistician)	33

FINDINGS – Q26 PLEASE PROVIDE YOUR NAME AND EMAIL IF YOU ARE INTERESTED IN (1) CONTRIBUTING TO DATA MANAGEMENT USE CASES, (2) PARTNERING WITH LIBRARIES AND RESEARCH COMPUTING, AND (3) PARTICIPATING IN MORE IN-DEPTH DATA MANAGEMENT ASSESSMENTS/EDUCATION/INTERVIEWS TO IMPROVE RESEARCH DATA MANAGEMENT AT UF.

Comments

"I teach data management and reproducible research approaches with open-source software."

"Thank you for the opportunity to contribute."

"I am interested in participating in more in-depth data management education and collaboration with Libraries and Research Computing."

"Interested in building research computing for international community in remote sensing."

"It is excellent that UF Libraries and HPC are partnering."

FUTURE PLANS

1. Develop general data management learning experience modules
 - Data Management Instructional Design intern (fall 2017 – spring 2018)
2. Develop relevant and discipline-specific data management use cases
 - Contact the thirty-six survey participants that answered Q26
3. Conduct 1st Annual Data Symposium – Enabling Data Reproducibility and Sustainability on March 19, 2018
 - <http://cms.uflib.ufl.edu/envisioning-data-symposium/Index.aspx>

Thank you

Questions/Comments

Contact information:

- Data Management and Curation Working Group (DMCWG), DATAMGMT-[at]LISTS[dot]UFL[dot]EDU
- Plato Smith, Data Management Librarian, plato[dot]smith[at]ufl[dot]edu
- Jean Bossart, Engineering Librarian, jean[dot]Bossart[at]ufl[dot]edu



EXPAND BEHAVIORAL SCIENCE RESEARCH WITH THE PREMIER RESOURCE FOR GRAY LITERATURE

Devoted to curating and indexing hard-to-find content from authoritative sources, PsycEXTRA® allows researchers to go beyond traditional peer-reviewed research materials. This high-quality and relevant database combines bibliographic records with unique full-text materials, focusing on the latest conference presentations and papers, newsletters, reports, patient-oriented factsheets and brochures, magazines, monographs, and standards and guidelines relevant to the needs of students, faculty, and clinicians alike.



Explore original, cutting-edge, high-quality research

- ▶ Updated biweekly
- ▶ More than 300,000 records and growing
- ▶ No coverage overlap with PsycINFO®, creating an ideal companion database
- ▶ Full text for more than 70% of records
- ▶ Ongoing updates of select professional literature from multiple state and government entities, associations, foundations, and more
- ▶ Average of 25,000 records or more added each year

Available via
EBSCO

Social Media Matters: Showing Up Online as Well as Ontime

Julia M. Gelfand, University of California, Irvine Libraries

Anthony Lin, Irvine Valley College Library, USA

Celebrating twenty-five years of acquiring, organizing and preserving grey literature in light of all the changes in publishing and information distribution gives us reason to highlight and parse communication technologies and methods. The organizational structure, emphasis on visualization and data, new applications of artificial intelligence, instant communication, and the convergence of technique with practice offers a lens in which to assess how social media has evolved. This suggests how opinions from certain sources matter as does the reshaping of consumer preferences that often dictate unexpected outcomes. Applications for business transactions, crowdsourcing, social learning, community building and flash mobs each attract different populations motivated by a range of goals. Matching these goals with appropriate social media conduits is the focus our work.

Social media today has demonstrated that eMail has given way to shorter and faster methods of communication such as text-messaging, tweeting and using any number of mobile apps to distill and manage one's inbox. The changes in communication strategies indicate that the level of expectations is reduced and challenges the retention of these transcripts for future needs. Generational differences and practices are studied by social and information scientists interested in the implication of how data elements in social media have proliferated and beg the question of whether they need to be managed, archived and preserved by whom and for how long. We also explore the changing values of social media. In this milieu of data reuse and repurposing, the multiple channels of social media are challenged to determine maximum efficiencies for alternative solutions in business and daily life. The unintended consequences of social media misuse are increasingly concerning in the wake of false reporting leading to fake news in our perceptions of creating a more evidence-based and transparent society.

Whether social media drives marketing or vice versa is ultimately what determines the prominence of greyness remains central to these queries. Methods of engagement and interaction vary depending on the form of social media and the capacity to channel different data points such as images and photos, opinions, preferences, promoting a brand or self-analysis. This paper explores how "the reach" to target audiences has become central to the goals of accomplishing business and social transactions and what the implications are for long term learning, commerce and communication effectiveness, a sustained reliance on social media and what the significance is for retaining trusted content in a variety of environments that currently emphasize living in the moment.

Bionotes

Julia Gelfand is the Applied Sciences, Engineering & Public Health Librarian at the University of California, Irvine Libraries where over the last 35 years she has performed many roles. She is active professionally and currently is a member of the Board of Directors of the Association of College & Research Libraries (ACRL), a division of the American Library Association, a member of the Science & Technology Section of the International Federation of Library Associations (IFLA) and is Secretary of Section T of the American Association for the Advancement of Science (AAAS). She writes and presents frequently on topics related to Scholarly Communication, Collection Management, Digital Scholarship, integration of multimedia in scientific literature, grey literature, social media, library as publisher. A graduate of Goucher College with graduate degrees from Case Western Reserve University, she is the recipient of many awards including the first GreyNet Award presented in 1999 and has been a Fulbright Fellow and a Thomas J. Watson Fellow. Email: jgelfand@uci.edu



Anthony Lin is the Head of Instruction and Collections at the Irvine Valley College Library. He holds a MSI from the University of Michigan-Ann Arbor, a BA in Spanish from California State University San Marcos, and a BS in Finance from San Diego State University. His interests are emerging technologies, effective bibliographic instruction, and collections management. Email: alin@ivc.edu





SOCIAL MEDIA MATTERS: SHOWING UP ONLINE AS WELL AS ONTIME

Julia Gelfand
Applied Sciences, Engineering &
Public Health Librarian
University of California, Irvine, USA
jgelfand@uci.edu

Anthony (Tony) Lin
Coordinator of Reference & Collections
Irvine Valley College, Irvine, CA, USA
alln@ivc.edu

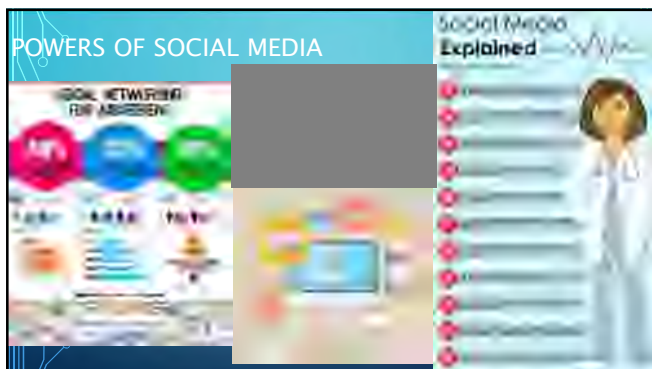
19TH INTERNATIONAL CONFERENCE ON GREY LITERATURE
ROME, ITALY
OCTOBER 23-24, 2017

INTRODUCTION

- Definition
- Chronology
- Social + Media + Networking
- Mobile Device Internet Usage Outpacing All Other Mediums
- Power & Reach of the Smartphone
- Engagement & Interaction

NEW LANDSCAPE







CONDUCTING LIFE DIFFERENTLY

- New dependencies
 - Science – fact vs myth
 - Creative entrepreneurship
 - Self empowerment
- Fewer boundaries
- Less Reliance on traditional news – identified three primary types of motives for news sharing on social media: altruistic, social, and self-serving
- Crowdsourcing, rating, reviews



RECENT EMERGENCY RESPONSIVENESS – DIFFERENT EXAMPLES FROM AROUND THE WORLD

HURRICANES 1,2,3



EARTHQUAKES, VOLCANIC
ERUPTIONS



COMFORTED BY TWEETS & POSTINGS...COMMUNICATIONS FROM HURRICANES HARVEY, IRMA, AND MARIA





LESSONS LEARNED – SOCIAL MEDIA MATTERS WHEN TIMES GET TOUGH

- Eliminates geographical barriers
- No time delay in transmission
- Simultaneous translation
- Serves all – demographically agnostic
- Identifies status of missing persons
- Tracking aid on the way
- Reliance on power to allow connectivity – charge phones; obtain cash from ATMs

STILL GREY & WHAT MATTERS NOW?

- Social media creates more expectations, thus change is new normal
- Smartphone optimization with integrated technologies is key to success of emergency responsiveness & bridges generation gaps
- Communication matters unifying family, bonding communities
- Television & news media rely upon social media from the public for frontline feeds
- Government services incorporating social media in delivery devoid of politics
- Challenges remain in organizing, retaining and managing social media output



QUESTIONS?

Video is the new Grey

Bastian Drees, Margret Plank, and Oleg Nekhayenko,
National Library of Science and Technology, Germany

Conference reports and conference proceedings are a crucial source of information as they document the current state of research in a scientific community. Moreover, these documents are mostly grey literature. However, in addition to printed conference proceedings, it is becoming more and more common practice to record and subsequently publish conference talks. As such, these videos are also an important element of contemporary scientific outputs and thus part of the cultural heritage. However, no sustainable standard has yet been established for handling these documents. Essentially, all of these videos are either published on commercial platforms like YouTube or Vimeo, on the conference webpage or not at all. Therefore, they are truly grey material.

While libraries have well-established procedures for collecting textual conference reports as part of the grey literature, comparable procedures for audio-visual conference recordings have not yet been established. On the other hand, according to the Meeting & EventBarometer 2016, more than half of the hosts and organizers express a need for action in setting up virtual platforms that complement the real life event.

Against this backdrop, we conducted an analysis of the needs and demands of conference hosts, organizers and service providers regarding audiovisual recordings. Qualitative interviews were conducted among 30 respondents to find out how widespread conference recordings are in engineering and the natural sciences. Furthermore, we wanted to obtain information on the use of these materials and about potential needs for support.

In this paper we present the results of the interviews. Almost half (47%) of the respondents stated that they already produced conference recordings (40%) or are planning to do so in the future (7%). The most common publishing platforms are the conference website (92%) and YouTube (46%). The most important aspects during video production and publication are rapid, cost-efficient and simple procedures, while after publication visibility and sustainability of the material are the most important concerns. The results of this study will be used to improve the workflows and services offered in the TIB AV-Portal, a web-based platform for quality-tested scientific videos such as conference recordings.

Bionotes

Bastian Drees works in the Competence Centre for non-textual Materials in the Research and Development Department of the German National Library of Science and Technology (TIB). He holds a PhD in physics and did his traineeship as a librarian at TIB and the Bavarian State Library (BSB) from 2014-2016. Email: Bastian.Drees@tib.eu



Margret Plank is currently the Head of the Competence Centre for Non-Textual Materials at the German National Library of Science and Technology in Hannover (Germany). The aim of the Competence Centre for Non-Textual Materials is to develop emerging tools and services that actively support users in the scientific work process enabling non-textual material such as audiovisual media, 3D objects and research data to be published, found and made available on a permanent basis as easily as textual documents. Previously she was responsible for Information Competence and Usability at the TIB. She has also worked as a researcher at the Institute of Information Studies and Language Technology at the University of Hildesheim. She represents TIB on a number of boards including IFLA Steering Committee Audiovisual and Multimedia Section as well as ICSTI / ITOC. Margret Plank holds a Master degree in information science and media studies from the University of Hildesheim, Germany. Email: margret.plank@tib.uni-hannover.de



Oleg Nekhayenko is research assistant at the German National Library of Science and Technology in the department of digital preservation. He received a bachelor's degree in Information Science and Language Technologies from Heinrich-Heine University Duesseldorf and a master's degree in International Information Management from University of Hildesheim. Email: Oleg.Nekhayenko@tib.eu



Impact of Emerging Information Technologies on Grey Literature

Dr. Dobrica Savić

Nuclear Information Section, International Atomic Energy Agency, NIS-IAEA, Austria

This paper reviews the emerging information technology (IT) trends and their potential impact on grey literature. Analysis will be done based on reports related to 2017 and issued by five leading international consulting and service companies with a special interest in researching the impact of IT on business environments, work procedures and behaviours. In addition, public awareness of the published reports expressed through pertinent comments left on the LinkedIn social network will be examined.

After reviewing the IT trends elaborated in the reports, as well as the LinkedIn comments, an attempt will be made to identify and aggregate the common trends that have the highest potential for defining and impacting the digital future of information management, more specifically that of grey literature. Each of the most common IT trends identified will be studied and elaborated individually, taking into consideration its direct impact on the creation, processing, dissemination and use of grey literature.

Information technology has become a driving force for change, innovation and improvements by offering powerful tools to further the digitization of our work processes and the services we offer, as well as the creation, distribution and use of information products such as grey and other literature. The emerging new digital information environment has the potential to affect all aspects of our work, the way we relate to our clients and customers, and the world around us. In order to meet expectations and benefit from this challenging opportunity, information managers need to be cognizant of new IT trends and the possibilities they offer in order to define the best strategies and action plans for their successful implementation.

Bionote

Dr. Dobrica Savić is Head of the Nuclear Information Section (NIS) of the IAEA. He holds a PhD degree from Middlesex University in London, an MPhil degree in Library and Information Science from Loughborough University, UK, an MA in International Relations from the University of Belgrade, Serbia, as well as a Graduate Diploma in Public Administration, Concordia University, Montreal, Canada. He has extensive experience in the management and operations of web, library, information and knowledge management, as well as records management and archives services across various United Nations Agencies, including UNV, UNESCO, World Bank, ICAO, and the IAEA. His main interests are creativity, innovation and use of information technology in library and information services. Email: d.savic@iaea.org



19th International Conference on Grey Literature
Public Awareness and Access to Grey Literature
23-24 October 2017, Rome

Impact of Emerging Information Technologies on Grey Literature

Dr. Dobrica Savić
d.savic@iaea.org

Nuclear Information Section
IAEA, Vienna

Presentation at a glance

- Grey literature definition
- Current grey literature challenges
- Review of information technology trends
- Trends impacting grey literature
 - AI and machine learning
 - Virtual and augmented reality
 - Internet of things
 - Digital platforms
 - Big data
 - Analytics
- Conclusions

Grey literature definition

Definition

Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body. ("Prague Definition" 2010)

The diverse and heterogeneous body of material that is made public outside, and not subject to, traditional academic peer review processes. (Adams *et al.* 2016)

Easier to describe than to define!



Grey literature definition

Multiple shades of grey

(a) Bibliographies Discussion papers Newsletters Unofficial press releases Program evaluation reports Technical notes Publications from governmental agencies Reports to funding agencies Unpublished reports Dissertations Policy documents	35 Reported manuscripts Unpublished manuscripts Conference abstracts Book chapters Research correspondence Scholarships Informal communications Census data Unpublished Seminars Plenary Workshops	Publications from National Commission on Violence WWW articles E-books Blogs and social media Data sets Commission reports Working papers Emergency releases Catalogues Speeches Reports on research
--	--	---

Data sets

Internet of Everything (IoE)
 Internet of Things (IoT)
 Industrial Internet of Things (IIoT)
 Machine to Machine communication (M2M)
 Self-driven cars
 Robots, sensors, security systems...

Estimates for the number of connected devices vary in billions. Gartner says some 20 billion by 2020. Allied Business Intelligence says more than 30 billion, Nelson Research says 100 billion, Intel says 200 billion, and International Data Co. says 212 billion.

Grey literature definition

Definition challenge

Due to multiple originators, volume, type and speed of GL creation, the focus of GL definition needs to shift to quality, intellectual property, curation and sustainability. Otherwise, it risks becoming obsolete due to its inability to differentiate GL from other documents.



New definition

GL is any recorded, referable and sustainable data or information resource of current or future value, made publically available without a traditional peer-review process.



Review of information technology trends

Gartner	Forbes	Forrester	Deloitte	Accenture
1. AI and advanced machine learning 2. Intelligent apps 3. Intelligent things 4. Virtual and augmented reality 5. Digital twin 6. Blockchain and Distributed Ledgers 7. Conversational System 8. Mesh App and Service Architecture 9. Digital Technology Platforms 10. Adaptive Security Architecture	1. IoT and smart home tech Web presence google Gartner Forbes Forrester Accenture Deloitte	Engagement technologies 1. IoT software and solutions 1740 144 9 1980 113	1. IT unbundled 2. Dark analytics Facebook 4080 4500 4 305 208	1. AI is the new UI 2. Ecosystems power play 3. Workforce marketplace 4. Design for humans 5. The uncharted
Gartner's Top 10 Strategic Technology Trends for 2017		1. Security automation and orchestration 2. Containers and container management 3. Edge computing 4. Cloud native application platforms 5. Hybrid wireless The Top Technology Trends to Watch: 2017 To 2021	6. Careers and learning 7. The employee experience 8. Leadership disrupted 9. People analytics 10. The future of work Global Human Capital Trends 2017: Rewriting the rules for the digital age	Technology Vision 2017 – Technology for People: The Era of the Intelligent Enterprise

Main information technology trends

- AI and machine learning
- Virtual and augmented reality
- Internet of things
- Digital platforms
- Big data
- Analytics

Disruptors!

Artificial intelligence and machine learning

- AI – systems that can think and act rationally like humans
- Very complex for development, maintenance and deployment
- Combine many technologies and techniques (e.g., deep learning, neural networks, natural-language processing (NLP))
- Move beyond traditional rule-based algorithms to create systems that understand, learn, predict, adapt and potentially operate autonomously
- Built into physical devices (e.g., robots, cars, consumer electronics, security), apps and services (e.g., virtual personal assistants, smart advisors, voice recognition, computer vision, translation, finance)
- AI becomes new user interface

Virtual and augmented reality

- VR takes us out of our reality and brings us to some other place
- AR takes our current reality and adds something to it
- VR vs. AR; scuba diving vs. going to the aquarium
- Virtual reality can bring us to a construction site where we can walk in any direction and see every detail
- Augmented reality is helpful for a client who can't visualize something. The idea is that a designer, an architect and a homeowner could sit around a table and look at the same 3D model on the table instead of a 2D plan. Human mind not able to tell the difference between computer-generated images and the real world
- Applications: military, medical, science, manufacturing, real estate, fashion, navigation, sightseeing, advertising and promotion, games like Pokémon Go

Internet of things

- The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction
- Requirements: things, unique identifier (IP), Wi-Fi, sensors, electronic circuits for control
- A 'thing' can be goods, objects, machines, appliances, buildings, vehicles, animals, people, plants, soil. E.g.:
 - a person with a heart monitor implant
 - a farm animal with a biochip transponder
 - an automobile with built-in sensors to alert the driver when tire pressure is low
 - connect with and learn about food, monitor supplies, search/locate, manage cities, control use of electricity, game immersion
- A move from people to computer-based data creation and capture
- Unimaginable complexity
- Privacy non existent
- Weapon of mass disruption

Digital platforms

A digital platform is a technology-enabled business model that creates value by facilitating exchanges between two or more interdependent groups.

The value increase as more members participate

b
e
n
e
f
i
t
s

Revenue - brings together end users and producers to transact with each other

Reduce cost - enables companies to share information in order to enhance collaboration or the innovation of new products and services

Collaboration - development, accelerated by third party application programming interfaces (APIs) enable participants to share data and create new services

Portability - cloud and other technologies provide resources on an as-a-service basis

Protection - intellectual property and data ownership protected to foster trust among participants

Advertising	Google, Baidu, Tencent, Redirect
Social	Facebook Twitter, Instagram, LinkedIn
Commerce	Amazon, Alibaba
Application stores	Apple App Store, Google Play
Crowd-sourcing	Uber, Blablacar, AirBnB
Enterprise resource planning (ERP)	SAP, Oracle, Infor

Market size value \$4.3 trillion (Accenture)

Big data

How big is BIG DATA?

Byte : one grain of rice

Kilobyte : one cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 semi-trucks

Terabyte : 2 container ships

Petabyte : Blankets Manhattan

Exabyte : Blankets west coast states

Zettabyte : Fills the Pacific Ocean

Yottabyte : AN EARTH SIZED RICE BALL!



Hobbyist



Desktop



Internet



Big Data

The Future?

Analytics

Analytics is the scientific process of transforming data into insight for making better decisions.

The Institute for Operations Research and the Management Science (INFORMS)

Business analytics explores past performance to gain insight and drive business planning.

The types include:

- **Descriptive** - provides simple summaries about the sample audience and about the observations made. Tell how things are going
- **Predictive** - anticipates what will happen, when and why it will happen. It uses statistical methods, but also machine learning algorithms, and heuristics, to extract information from data and predict trends and behavior patterns

Applications of predictive analytics

- Analytical customer relationship management (CRM)
- Clinical decision support systems
- Collection analytics
- Cross-sell
- Customer retention
- Direct marketing
- Fraud detection
- Portfolio, product or economy-level prediction
- Risk management
- Underwriting

Conclusions

Emerging environment

- Production moves to Industry 4.0
- Emerging new technologies
- Machines are learning to think
- Machines 'talk' to each other
- VR and AR experience
- Automatic creation and processing of massive data
- High level work specialization
- Extreme complexity
- Constant learning and development
- Liquid structure of info. & docs.
- Commercialization of information assets

Concept

- Make Distinction from other forms
- Cover new electronic forms

Processing

- Provide systematic collection
- Improve source reliability
- Quality bibliographic control
- Create standard key metadata

Sustainability

- Assure long-term preservation
- Persuade financial sustainability

Usability

- Protect intellectual property
- Provide open access
- Secure privacy

***Once a new technology rolls over you,
if you're not part of the steamroller,
you're part of the road.***

Stewart Brand

Thank you!

Apps & Codes: Making profiles for fluid publishing contents

Flavia Cancedda, CNR, National Research Council - Central Library, ISSN National Centre

Luisa De Biagi, CNR, National Research Council - Central Library, OpenGREY Network, Italy

Apps are one of the most used digital publishing tools and convey intellectual contents for innumerable functionalities. App softwares present technical characteristics different from each other, depending on the intellectual content for which they are profiled and made available. In publishing field, apps that convey traditional editorial products disclose the maximum of their innovative potential in representing the interface of continuously updating products (such as newsletters, magazines, newspapers, guides and tourist maps, blogs, open/e-gov/research data systems and related data-bases, clinical trials, forums and websites with specific matter or content, etc.). If apps are the interface software – thus, the communication interface - of publishing products, could it be possible to make them identifiable (and manageable) through the same bibliographic codes used for the corresponding traditional publishing products (e.g., ISBN, ISSN, DOI, etc.)? We think that when apps show and preserve the essential bibliographic identifying data, the answer could be no less than positive.

The goal of this study is to show, through a comparative analysis of National and European case-studies and best practices examples, how the potential public usefulness (not only commercial but also informative) of a bibliographic identification for apps is similar to the usefulness recognized in case of traditional publishing products, opening new scenarios and results also for grey products: a new traceability skill could be established for apps containing permanent references, traditionally considered necessary for the identification of editorial objects (title, publisher, year, updating mode or frequency, etc.); traceability through numerical codes would allow also easier dissemination, marketing, indexing processes by search engines, portals and sales store, up to make indexing tools for bibliographic services and librarians more specific and professional.

Bionote

Flavia Cancedda, Technologist, librarian and information specialist. Head of the Italian ISSN Centre (Central Library CNR); Component of “Documentation and Information” Committee of the UNI, National standardization organism (Chairman of the Subcommittee “Presentation, identification and description of documents”; component of the Working Group on National Standard about Professional Requirements for Librarians); Component of the ACNP Libraries Committee (Archivio Collettivo Nazionale Periodici) and of the ACNP Cataloguing Rules Revision Working Group. Since 1991 to 2000 librarian in two different Italian universities and in the National Library of Florence (section: National Bibliography); since 2001 librarian in the Central Library of the Italian National Research Council (CNR). Master’s degree in Libraries Management and Direction. Post-graduate degree as Manuscripts conservator at the Special School for Archivists and Librarians (La Sapienza University in Rome). Author of books and essays about the history of libraries and historical bibliography.

Email: flavia.cancedda@cnr.it

Luisa De Biagi got her Laurea Degree in Literature and Philosophy at 'La Sapienza' Univ. of Rome (Art history and Cultural heritage). With a Specialization in 'Archivist-Palaeographer' (Vatican School of Palaeography, Diplomatics and Archivistics at the Vatican Secret Archive) as well as a Specialization Degree in Archivistics, Palaeography and Diplomatics (Archivio di Stato, Rome) and a Degree from the Vatican School of Library Sciences. De Biagi further holds a Master in 'Business Publishing' (LUISS Management – Rome). She’s been working for the SIGLE Network (System for Information on Grey Literature in Europe) since 2002, Since 2010 she’s is responsible for the Italian National Referring Centre of Grey Literature at CNR Central Library ‘G. Marconi’ as representative of the European Network and Openarchive OpenGrey (System for Information on Grey Literature in Europe). She’s taken part in 3 editions of the Annual International Congress on Grey Literature – GL (GL5, Amsterdam, GL13, Washington d.c. GL14, Rome and GL15 at Bratislava). She’s is also a member of the CNR Working Group for Cedefop-Refernet Project (Consortium for Professional Education and Training coordinated by ISFOL), the Committee for Legal Deposit Acquisition at CNR Central Library, and a member of the European Association of Health Information and Libraries (EAHIL). She’s also responsible for the Library Functional Units ‘Education and Training’ and ‘Cultural Activities Management’, organizing didactics laboratories for students, professional training courses and teaching in professional trainings for librarians, students and users. Email: luisa.debiagi@cnr.it

GL 19 – XIX International Conference on Grey Literature –
National Research Council of Italy
Roma (IT), October 23-24, 2017

**Apps & Codes: Making Profiles
for Fluid Publishing Contents**

by Flavia Cancedda* & Luisa De Biagi* **

CNR – National Research Council, Rome (IT)
Central Library 'G. Marconi'
* National ISSN Centre** National Centre for Grey Literature



Apps are one of the most used digital publishing tools...

... and convey intellectual contents for innumerable functionalities.
App programs present technical characteristics different from each other,
depending on the intellectual content for which they are profiled and made
available

In publishing field the maximum innovative potential of apps for continuously updating products

Apps that convey traditional editorial products disclose the maximum of their
innovative potential in representing the interface of continuously updating
products
such as
*newsletters, magazines, newspapers, guides and tourist maps, blogs, open/e-
gov/research data systems and related data-bases, clinical trials, forums and
websites with specific matter or content, etc.*

Question:

If apps are the interface software – thus, the
communication interface – of publishing
products, could it be possible to make them
identifiable (and manageable) through the
same bibliographic codes used for the
corresponding traditional publishing
products (e.g.: ISBN, ISSN, DOI ...)?

and positive answer:

We think that when apps show and preserve
the essential bibliographic identifying data,
the answer could be no less than positive.



positive answers from ISBN/ISSN world

In ISBN Manual, par. 6.1/6.2: *Where a publication is available electronically (e.g., an e-book, e-book app, CD-ROM, or publication available on the Internet), it will qualify for an ISBN provided that it contains text and is made available to the public ... An ISBN may be used to identify a specific software product that is intended for educational and/or instructional purposes, such as a computer-based training product, provided that it is neither customizable nor requires data in order to function"*

ISSN ISO standard (3297:2007): *ISSN can be assigned for ongoing integrating resource: "continuing resource that is added to or changed by means of updates that do not remain discrete and are integrated into the whole"*

positive answers also from DOI® world

<https://www.doi.org/factsheets/DOIKeyFacts.html>

DOI® is applicable to any object (= any entity or thing: physical, digital, or abstract; resources, parties, licences, etc.)

- › is a digital Identifier of an object = network actionable identifier ("click on it and do something")
- › Initial focus on entities was documents/media e.g., articles, data sets; now moving into parties, licenses and other sectors

DOI® provides an actionable, interoperable, persistent link

- › Actionable – through use of identifier syntax and network resolution mechanism (Handle System®)
- › Persistent – through combination of supporting improved handle infrastructure (registry, proxy, etc.) and social infrastructure (obligations by Registration Agencies)
- › Interoperable – through use of a data model providing semantic interoperability and grouping mechanisms"

public usefulness in identifying apps

What public usefulness of a bibliographic identification for apps?

1. Bibliographic (as for traditional publishing products: **indexing, searching, individuating** etc.)
2. Informative (for all kind of publishing informational transaction, including **copyright field** and related rights of the owners)
3. Commercial (for all kind of **"financial" publishing transaction**: sales, subscriptions, discounts, fees, duties, royalties, taxes

New traceability skill...

A new traceability skill could be established for apps containing permanent references, traditionally considered necessary for the identification of editorial objects (title, publisher, year, updating mode or frequency, etc.); by identifying codes traceability through numerical codes (as ISBN, ISSN, DOI or others) would allow also easier dissemination, marketing, indexing processes by search engines, portals and sales store, up to make indexing tools for bibliographic services and librarians more specific and professional

Best practices and DOI applications/trends

- › QR Code Generator by Cross Ref :
- › <https://www.crossref.org/labs/qr-code-generator/>: The US DOI Agency is experimenting a new qr-encoded Crossref DOI, inspired by Google recent promotion of QR codes: you could generate a QR Code for any given Crossref DOI, even postcards or media gadget

Best practices and DOI trends: Github and Zenodo

- › Digital Object Identifiers (DOI) are fundamental in academic references and metrics system. For software open developpers: Githubs
→ <https://github.com/open-sources>, ann opensource an open source platform for writing, sharing and creating software citable by archiving on GitHub repositories and assigning a DOI with the data archiving tool Zenodo <https://zenodo.org/>

Best practices and DOI trends: Github and Zenodo

- Zenodo <https://zenodo.org/>
- Name inspired by Zenodoto, 1st director of Alessandria Library
- Zenodo integrates with GitHub “to make code hosted in GitHub citable”
- European Open-data set multidisciplinary repository by CERN and OpenAIRE European Project
- DOI assigned to every data-set
- ZENODO interacts also with Mendeley (recently by acquired by Elsevier), ORCID and ResearchID

References

- ISO 2108:2005, International standard book number (ISBN)
- ISO 3297:2007, International standard serial number (ISSN)
- ISO 26324:2012, Digital object identifier system [DOI]
- ISO 17316:2015, International standard link identifier (SLI)
- Other past Internet standards:*
- RFC 1630 (1994), <https://tools.ietf.org/html/rfc1630>
- RFC 2141 (1997), <https://tools.ietf.org/html/rfc2141>
- RFC 3986 (2005), <https://tools.ietf.org/html/rfc3986>
- RFC 3987 (2005), <https://tools.ietf.org/html/rfc3987>
- Making Your Code Citable, Github guides
<https://guides.github.com/activities/citable-code/>

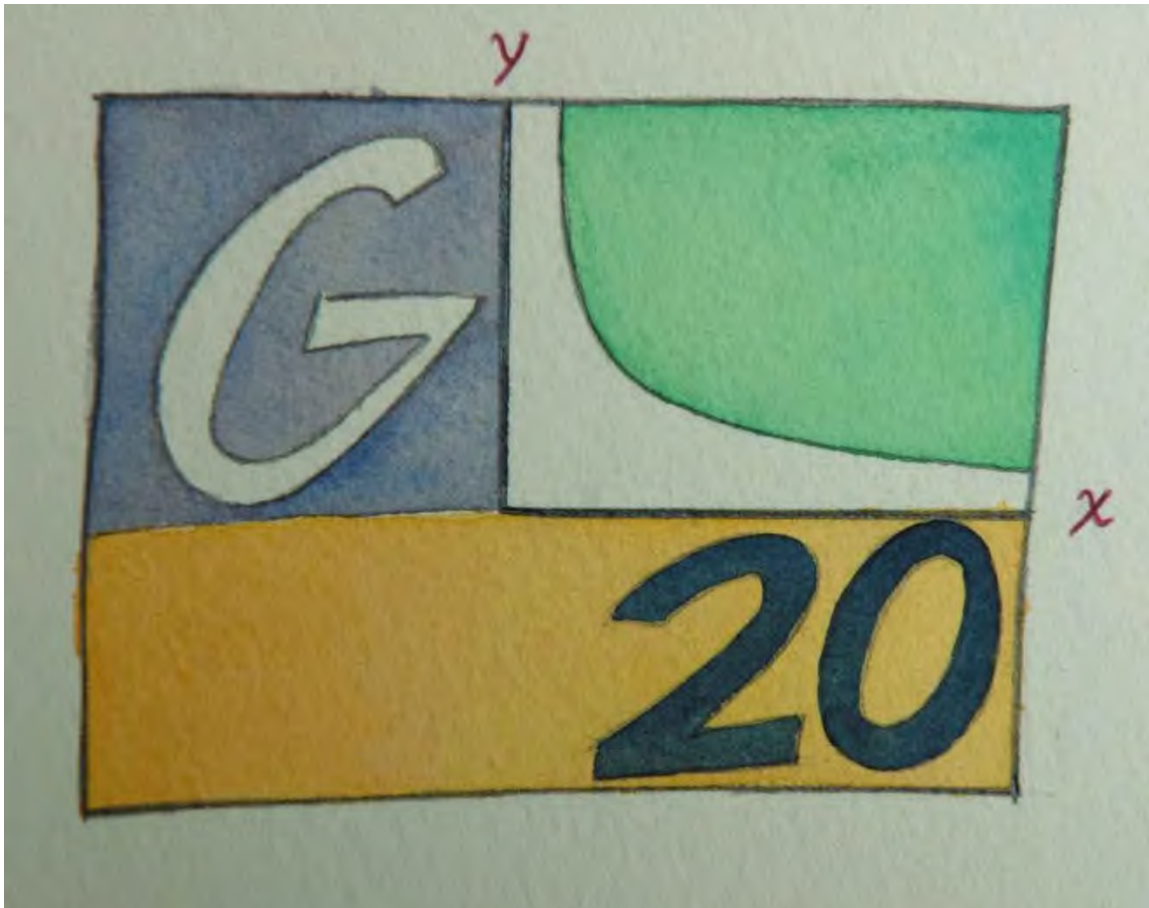
List of Participating Organizations

As of October 9, 2017

Access Innovations, Inc.	United States
ACM Digital Library	United States
Alberta Health Services	Canada
Biblioteca Centrale, G. Marconi; CNR	Italy
Centre National de Recherche Scientifique, CNRS	France
Data Archiving and Networked Services, DANS-KNAW	Netherlands
Deanet Media Company	Italy
EBSCO	United States
EiPass	Italy
Federal Library Information Network, FEDLINK	United States
Food and Agriculture Organization of the United Nations, FAO	Italy
Georgetown University	United States
GERiiCO laboratory	France
German National Library of Science and Technology, TIB	Germany
Grey Literature Network Service, GreyNet International	Netherlands
Institute for Applied Mathematics and Information Technologies, IMATI-CNR	Italy
Institut de l'Information Scientifique et Technique, Inist-CNRS	France
Institute of Clinical Physiology, IFC-CNR	Italy
Institute of Computational Linguistics, ILC-CNR	Italy
Institute of Informatics and Telematics, National Research Council	Italy
Institute of Information Science and Technologies, ISTI-CNR	Italy
Irvine Valley College	United States
Istituto Motori; IM-CNR	Italy
Korea Institute of Science & Technology Information, KISTI	Korea
Kumoh National Institute of Technology	Korea
Library of Congress, LoC	United States
National Civil Service Project, UNSC	Italy
National Library of Technology, NTK	Czech Republic
National Research Council of Italy, CNR Central Library	Italy
New York Academy of Medicine	United States
New York University; School of Medicine	United States
Nuclear Information Section; International Atomic Energy Agency, NIS-IAEA	Austria
Pratt Institute, School of Information	United States
PricewaterhouseCoopers, PwC	Netherlands
PUSAN National University	Korea
Saint Petersburg State University	Russia
Slovak Centre of Scientific and Technical Information, CVTISR	Slovakia
Springer Nature	Italy
TextRelease, Program and Conference Bureau	Netherlands
University of Arizona, College of Medicine Phoenix	United States
University of British Columbia, UBC	Canada
University of California, Irvine Libraries, UCI	United States
University of Florida; George A. Smathers Libraries	United States
University of Ghent, Heymans Institute of Pharmacology	Belgium
University of Liège, Department of General Practice	Belgium
University of Lille 3	France
University of Maryland; University Libraries	United States
University of Rouen, Department of Information and Medical Informatics	France
University of Texas Health Science Center at Houston, School of Biomedical Informatics	United States
University of Toronto, Institute of Health Policy, Management and Evaluation	Canada
University of Wisconsin, Milwaukee	United States
Wikimedia	Czech Republic
Wiley	Italy
WorldWide ScienceAlliance	United States

Twentieth International Conference on Grey Literature

Research Data Fuels and Sustains Grey Literature



Loyola University
St. Charles Avenue
New Orleans, Louisiana USA
December 3-4, 2018



In Memory of the Late Mary Lorraine Farace, A Native of New Orleans in Heart and Soul

'Public Awareness and Access to Grey literature'

National Research Council of Italy, CNR Rome * October 23-24, 2017

Publication Order Form

NINETEENTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

Publication(s):	No. of Copies	x	Amount in Euros	Subtotal
GL19 CONFERENCE PROCEEDINGS - Printed Edition ISBN 978-90-77484-31-9 ISSN 1386-2316 <i>Postage and Handling excluded*)</i>		x	109.00 = €	
GL19 CONFERENCE PROCEEDINGS - PDF Edition ISBN 978-90-77484-31-9 ISSN 1386-2316 <i>Forwarded via email</i>		x	109.00 = €	
GL19 Conference Proceedings - Online Edition ISBN 978-90-77484-31-9 ISSN 2211-7199 <i>Password Protected Access</i>		x	109.00 = €	



*POSTAGE AND HANDLING PER PRINTED COPY *)*

Holland		x	5.00	€
Other		x	15.00	€
TOTAL EURO =				€

Customer Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
E-mail Address:	

☐ Direct transfer to TextRelease, Rabobank Amsterdam
BIC: RABONL2U IBAN: NL70 RABO 0313 5853 42, with reference to "GL19 Publication Order"

☐ MasterCard/Eurocard ☐ Visa Card ☐ American Express

Card No. _____ Expiration Date: _____

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

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

Place: _____ Date: _____

NOTE: CREDIT CARD TRANSACTIONS WILL BE AUTHORIZED VIA OGONE/INGENICO DESIGNATED PAYMENT SERVICES

TextRelease
www.textrelease.com

GL19 Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: info@textrelease.com

Index to Authors

A-B

Balashova, Yuliya B.	31
Barsotti, Alessia	77
Bartolini, Roberto	60, 63
Biagioni, Stefania	17, 77
Bossart, Jean	97

C

Cancedda, Flavia	115
Candela, Leonardo	79
Cardillo, Elena	49
Carlesi, Carlo	77
Carlson, Kathleen Noma	70
Černohlávková, Petra	69
Chamberlain Kritikos, Katie	37
Cho, Hwan Gue	73
Choi, Ki Seok	73
Choi, Kwang-Nam	73,75

D-E

Darmoni, Stefan	49
De Biagi, Luisa	115
Deluca, Rosaria	17
Doorn, Peter	91
Dorris, C. Scott	23
Drees, Bastian	108

F-G

Farace, Dominic	25, 63
Frantzen, Jerry	25, 63
Frosini, Luca	79
Gelfand, Julia M.	103
Giannini, Silvia	17
Giustini, Dean	64
Goggi, Sara	60, 63
Grosjean, Julien	49

H-I-J

Hlava, Marjorie M.K.	24
Ittoo, Ashwin	49
Jamouille, Marc	49
Jung, Yu-Chul	75

K-L

Kritikos, Katie Chamberlain	37
Lim, Chul-Su	73, 75
Lin, Anthony	103
Lipinski, Tomas A.	37

M-N

Mack, Daniel C.	43
Molino, Anna	17
Monachini, Monica	60
Nekhayenko, Oleg	55, 108

P-R

Pagano, Pasquale	79
Pardelli, Gabriella	60, 63
Pardini, Andrea	77
Plank, Margret	108
Prost, Hélène	85
Resnick, Melissa P.	49
Russell, Judith C.	24
Russo, Irene	60, 63

S-T-U

Savić, Dobrica	109
Schöpfel, Joachim	70, 85
Shin, YongJu	73
Smith, Plato L.	25, 97
Son, Jiseong	73, 75
Subirats, Imma	15

V-Z

Vander Stichele, Robert	49
Vanmeerbeek, Marc	49
Vaska, Marcus	64, 70
Vyčítalová, Hana	64, 69