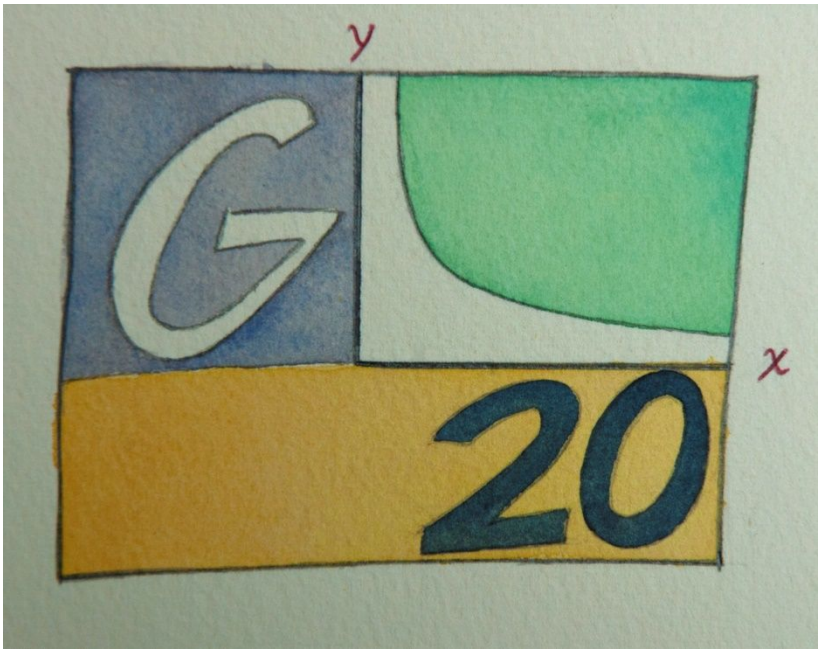
Twentieth International Conference on Grey Literature

Research Data Fuels and Sustains Grey Literature

Loyola University New Orleans, USA • December 3-4, 2018

Program Book

ISSN 1385-2308



Program and Conference Sponsors

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

CENTRUM VEDECKO-TECHNICKÝCH INFORMACIÍ SR

DANS

Nuclear Information Section
NIS - IAEA
International Atomic Energy Agency

TIB LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY

KiSTi www.kisti.re.kr

Inist cnrs

EBSCO PUBLISHING

LIBRARY OF CONGRESS FEDLINK

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library

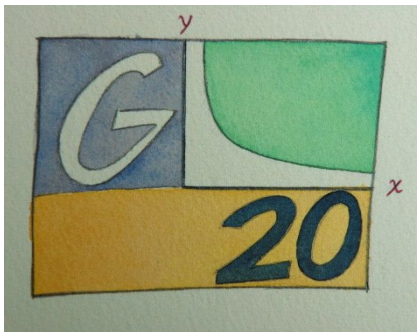
UF George A. Smathers Libraries UNIVERSITY of FLORIDA

UNIVERSITY of WISCONSIN UWMILWAUKEE School of Information Studies

GL20 Program and Conference Bureau

TextRelease

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



CIP

GL20 Program Book
Twentieth International Conference on Grey Literature “Research Data Fuels and Sustains Grey Literature”. - Loyola University New Orleans, USA December 3-4, 2018 / compiled by D. Farace and J. Frantzen ; GreyNet International, Grey Literature Network Service. – Amsterdam : TextRelease, December 2018. – 136 p. – Author Index. – (GL Conference Series, ISSN 1385-2308 ; No. 20).
TIB (DE), DANS-KNAW (NL), FEDLINK-Library of Congress (USA), CVTISR (SK), EBSCO (USA), Inist CNRS (FR), ISTI-CNR (IT), KISTI (KR), NIS-IAEA (UN), NTK (CZ), and University of Florida; George A. Smathers Libraries (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.

ISBN 978-90-77484-34-0

Foreword

RESEARCH DATA FUELS AND SUSTAINS GREY LITERATURE

The definition of research data is as encompassing as the field of grey literature. What should be included and what should be excluded is and remains an issue of concern. Research data can be defined as factual materials collected by diverse communities of practice required to validate findings. While the majority of research data is created in digital format, research data in other formats cannot be excluded. The formats in which research data appear are multiple and the types of research data are diverse. This also holds for the numerous document types in which grey literature appear published.

Today, while emphasis is placed on big data, the fact that the majority of research projects are small to medium size is overlooked. This is but another characteristic that holds true for grey literature. Nonetheless, one should be aware that research publications are not research data, for they are often managed separately from one another. Just as there are a number of stakeholders involved in the production, access, and preservation of grey publications, so too are there stakeholders tasked with the creation and management of research data. Libraries and data management librarians have the responsibility for the curation of the data they collect and preserve. And, it is important to stress the need to maintain appropriate metadata related to research data in order to facilitate their interpretation and further reuse.

Over the past quarter century, grey literature communities have worked diligently to demonstrate how their documents are produced, published, reviewed, indexed, accessed, and further used, applied, and preserved. Today, these communities are now challenged to demonstrate how research data fuels and sustains their grey literature. These communities of dedicated researchers and authors maintain a strong conviction in the uses and applications of grey literature for science and society. Through the years, they have proved willing to share the results of scholarly work well beyond their own institutions. Hence, one can assume they are aware that innovation forfeits with the loss of data as with the loss of information. This 20th International Conference in the GL-Series seeks to address key issues and topics related to grey literature and its underlying research data.

Dominic Farace — *Amsterdam,*
GREYNET INTERNATIONAL — DECEMBER 2018

GL20 Conference Sponsors

ISTI, Italy

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

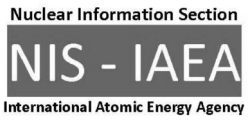
CVTISR, Slovak Republic

Slovak Centre of Scientific and Technical Information

KISTI, Korea

Korea Institute of Science and Technology 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

GL20 Conference Sponsors (CONTINUED)

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

UW-Milwaukee-SOIS, USA
School of Information Studies (SOIS)
University of Wisconsin, Milwaukee

UF, USA
George A. Smathers Libraries
University of Florida

GL20 Program Committee

Brian Hitson [Chair]
Office of Scientific and Technical Information; U.S. Department of Energy

Robert Bell
Loyola University New Orleans, USA

George Barnum
U.S. Government Publishing Office

Meg Tulloch
U.S. Government Accountability Office

Margret Plank
German National Library of Science and Technology, Germany

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

Silvia Giannini
Institute of Information Science and Technologies ISTI-CNR, Italy

Ján Turňa
Slovak Centre of Scientific and Technical Information Slovak Republic

Stefania Biagioni
NeMIS Research Laboratory
Italy

Joachim Schöpfel
University of Lille
France

Judith C. Russell
University of Florida Libraries
USA

Plato L. Smith
University of Florida; George A. Smathers Libraries, USA

Henk Harmsen
Data Archiving and Networked Services, Netherlands

Marcus Vaska
Alberta Health Services
Canada

Dominic Farace
GreyNet International
Netherlands

Tomas A. Lipinski
University of Wisconsin Milwaukee, USA

Table of Contents

Program Chair and Conference Moderators

Moderator Day One

Robert Bell
Director of Learning Resources
Loyola University New Orleans

Robert Bell is Director for the Office of Writing and Learning Services at Loyola University. He was born and raised in New Orleans and its confines. He received his education in New Orleans: B.A. from Loyola University New Orleans, M.F.A. from the University of New Orleans. In addition to being the Director for OWLS, he teaches undergraduate writing and literature courses, and directs the Loyola Ireland Summer Abroad program.

He recently published "America's Disaster Culture: The Production of Natural Disasters in Literature and Pop Culture" and edited "Eco Culture: Disaster, Narrative, Discourse," both with Robert M. Ficociello. Additionally, Bell is the Disasters, Apocalypses, and Catastrophes area Co-Chair for the Popular Culture/American Culture Association.

Email: rcbell@loyno.edu

Program Chairman

Brian A. Hitson
Director
U.S. Department of Energy; OSTI

Brian Hitson is Director of the U.S. Department of Energy (DOE) Office of Scientific and Technical Information (OSTI), a DOE corporate function that is managed by the Office of Science. OSTI fulfills agency-wide responsibilities to collect, preserve, and disseminate scientific and technical information emanating from DOE research and development (R&D) activities. Prior to becoming Director, Brian led a range of OSTI's programmatic and administrative activities. He has also managed OSTI's international information exchange programs, as well as the digitization and preservation of a scientific document repository. As Director, he has led strategic efforts to improve discoverability and linkages between diverse, related research objects, including publications, datasets, and scientific software available at OSTI.GOV. He played a key role in the development of WorldWideScience.org and in the establishment of the WorldWideScience Alliance in 2008.

Email: HitsonB@osti.gov

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, 3 December	DAY 2 - Tuesday, 4 December
Loyola University, Audubon Room, 6363 St. Charles Ave.	***Loyola University, Audubon Room, 6363 St. Charles Ave.***
Registration Desk Open 8:30-9:00 am	**Registration Desk Open 9:00-9:30 am**
Opening Session 9:00-10:30 am ***Welcome, Keynote Address, Opening Paper***	**Poster Session, Sponsor Showcase 9:30-11:00 am** ***Presentations and Exhibits***
Morning Break 10:30-11:00 am	
Session One 11:00-1:00 pm **Research Data and Open Access Compliance**	**Panel Session 11:00-12:30 pm** ***WorldWide Science Alliance*** **Global Science Gateway, 2008-2018**
Lunch 1:00-2:00 pm	**Lunch 12:30-1:30 pm**
Session Two 2:00-4:00 pm **Data Management and the Role of Librarians**	**Session Three 1:30-3:30 pm** **Current Research Trends in Grey Literature**
Afternoon Break 4:00 -4:30 pm	**Pause 3:30-3:45 pm**
Introduction to Posters 4:30-5:30 pm ***Poster Briefings***	**Closing Session 3:45-4:30 pm** ***Wrap-up, Poster Prize, Handoff, Farewell***
 **New Orleans Landmark**	CITY OF NEW ORLEANS **1718 - 2018**

DAY ONE

Moderator: Robert Bell, Loyola University New Orleans, United States

09:00 – 10:30 am

OPENING SESSION

Welcome Address, *Brian Hitson, Director Office of Scientific and Technical Information, OSTI-DOE, United States* **13**

Keynote Address, **A Roadmap for Enabling Integrated Science: The USGS Experience in Open Data** **14**
Vivian Hutchison, Acting USGS Library Director, U.S. Geological Survey, United States

Opening Address, **The U.S. Government Publishing Office:** **15**
Keeping America Informed in the 21st Century and Beyond
Cynthia Etkin, Senior Program Planning Specialist, GPO
U.S. Government Publishing Office, United States

10.30 -11:00 am *Morning Break*

11:00 – 1:00 pm

SESSION ONE – RESEARCH DATA AND OPEN ACCESS COMPLIANCE

When is 'grey' too 'grey'? A case of grey data **17**
Dobrica Savić, Nuclear Information Section, International Atomic Energy Agency, NIS-IAEA, Austria

Big data, grey literature and personal data protection. A crucial impact on social life **22**
Luisa De Biagi, CNR Central Library; Roberto Puccinelli, Network and Informative Systems Office; CNR,
Antonio Tintori, Institute for Research on Population and Social Policies, IRPPS-CNR, Italy

Legal Issues Surrounding the Collection, Use and Access to Grey Data in the University Setting; **28**
How Data Policies Reflect the Political Will of Organizations
Tomas A. Lipinski, School of Information Studies; University of Wisconsin-Milwaukee, United States
Kathrine A. Henderson; LibSource, A LAC-Group Company, United States

Creating, using, and distributing compliant research data: A model for fulfilling legal and policy requirements **34**
Daniel C. Mack; University of Maryland, United States

On Open Access to Research Data: A DANS Use Case **39**
Emilie Kraaikamp, Chris Baars, Henk Harmsen, and Hella Hollander, Data Archiving and Networked Services,
DANS-KNAW, Netherlands

1:00 - 2:00 pm *Lunch*

2:00 – 4:00 pm

SESSION TWO: DATA MANAGEMENT AND THE ROLE OF LIBRARIANS

The data librarian: myth, reality or utopia? **45**
Silvia Giannini and Anna Molino, Institute of Information Science and Technologies, ISTI-CNR, Italy

Research Data Management: What can librarians really help? **51**
Yuan Li, Willow Dressel and Denise Hersey; Princeton University, United States

Data Management and the Role of Librarians **57**
Plato L. Smith and Sara Gonzalez, George A. Smathers Libraries; University of Florida, United States

Measuring Reuse of Institutionally-Hosted Grey Literature **63**
Elizabeth Kelly, Loyola University New Orleans; Caroline Muglia, University of Southern California;
Genya O'Gara, Virtual Library of Virginia; Ayla Stein Kenfield, University of Illinois at Urbana–Champaign;
Santi Thompson, University of Houston; Liz Woolcott, Utah State University, United States

Librarians' Role in GAO Reports **69**
Meg Tulloch, Colleen Candrl, and Leia Dickerson; U.S. Government Accountability Office, GAO, United States

4:00 - 4:30 pm *Afternoon Break*

4:30 – 5:30 pm

INTRODUCTION TO CONFERENCE POSTERS AND SPONSOR SHOWCASE

Those presenting posters will have the opportunity to introduce their work in the Main Conference Room. The Poster Session and Sponsor Showcase will continue on the morning of **Day Two**, where the presenters will meet with conference participants in an informal setting. The Poster Prize 2018 will be presented during the Closing Session.

DAY TWO

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

09:30 – 11:00 am

POSTER SESSION AND SPONSOR SHOWCASE

Title of the Posters, Authors, and their Organizations **12**

11:00 – 12:30 pm

PANEL SESSION – WORLDWIDE SCIENCE*Alliance* AND GLOBAL SCIENCE GATEWAY, 2008-2018 **99**

Brian Hitson [Panel Moderator] Director, Office of Scientific and Technical Information; U.S. DOE, United States

Dobrica Savić, Nuclear Information Section; International Atomic Energy Agency, United Nations **100**

Lorrie Johnson, Office of Scientific and Technical Information; U.S. DOE, United States **104**

Abe Lederman, Deep Web Technologies, Inc., United States **108**

Justin Fessler, IBM Corporation, United States **112**

12:30 -1:30 pm *Lunch*

1:30 – 3:30 pm

SESSION THREE – CURRENT RESEARCH TRENDS IN GREY LITERATURE

A Case Study on the Content Curation for Improving Effectiveness of Research Reports **117**
Seokjong Lim and Heykyong Hwang; Korea Institute of Science & Technology Information, KISTI, South Korea

Open Data engages Citation and Reuse: A Follow-up Study on Enhanced Publication **119**
Dominic Farace, Jerry Frantzen, GreyNet International, Netherlands; Joachim Schöpfel, University of Lille, France

The Q-Codes: Metadata, Research data, and Desiderata, Oh My! Improving Access to Grey Literature **124**
Melissa P. Resnick, University of Texas Health Science Center at Houston, United States [et al]

Analysis of folk literature in grey literature from the National Library of China **125**
Cui Yue, National Library of China, Beijing, China

When the Virtual Becomes Reality: An Environmental Scan of the Presence of Virtual Reality and Artificial Intelligence in Health and Cancer Care Environments **126**
Marcus Vaska, Alberta Health Services; Knowledge Resource Service, Canada

3:30 – 3:45 pm *Afternoon Break*

3:45 – 4:30 pm

3:45 PM CLOSING SESSION – REPORT MODERATORS, POSTER PRIZE, CONFERENCE HANDOFF, FAREWELL

4:30 PM POST CONFERENCE TOUR: MONROE LIBRARY, LOYOLA UNIVERSITY NEW ORLEANS

LOYOLA UNIVERSITY NEW ORLEANS

Poster Session and Sponsor Showcase

Welcome Address

Brian A. Hitson
Director
U.S. Department of Energy; OSTI

Brian Hitson is Director of the U.S. Department of Energy (DOE) Office of Scientific and Technical Information (OSTI), a DOE corporate function that is managed by the Office of Science. OSTI fulfills agency-wide responsibilities to collect, preserve, and disseminate scientific and technical information emanating from DOE research and development (R&D) activities. Prior to becoming Director, Brian led a range of OSTI's programmatic and administrative activities. He has also managed OSTI's international information exchange programs, as well as the digitization and preservation of a scientific document repository. As Director, he has led strategic efforts to improve discoverability and linkages between diverse, related research objects, including publications, datasets, and scientific software available at OSTI.GOV. He played a key role in the development of WorldWideScience. org and in the establishment of the WorldWideScience Alliance in 2008. Email: HitsonB@osti.gov

Keynote Address

A Roadmap for Enabling Integrated Science: The USGS Experience in Open Data

Vivian Hutchison
Acting USGS Library Director,
U.S. Geological Survey, United States

Vivian (Viv) Hutchison is currently serving as Acting Director for the US Geological Survey Library. The Library was established in 1879 and is considered one of the largest Earth Science libraries in the world. Viv also holds the position of Branch Chief for the Science Data Management Branch in the Core Science Analytics, Synthesis and Library (CSASL) program in the US Geological Survey (USGS). Her team supports scientists in long-term management and release of scientific data through tools and best practices. She holds a Masters of Library Science (MLS) from the University of Maryland College Park, and lives in Denver, Colorado. Viv has worked for the USGS since 2002. Email: vhutchison@usgs.gov

Opening Address

The U.S. Government Publishing Office: Keeping America Informed in the 21st Century and Beyond

Cynthia Etkin, Senior Program Planning Specialist, GPO
U.S. Government Publishing Office, United States

On June 23, 1860, the 36th Congress of the United States approved a joint resolution that created the Government Printing Office (GPO), and directed the Superintendent of Public Printing to have executed the printing and binding of documents approved by the Senate and House of Representatives, and the executive and judicial branch departments. GPO began operation with 350 employees on the day that Abraham Lincoln was inaugurated as the 16th President of the United States, March 4, 1861. The GPO has a long history and it is a remarkable story.

The Printing Act of 1895 overhauled existing printing laws and created a number of new authorities for GPO. Among them was to disseminate Government public documents through three channels: deposit in designated libraries; distribution to Government entities; and sales. Another new authority was to organize the Government's information, with the provision of a cataloging and indexing program. From this landmark legislation came two important public information programs, the Cataloging and Indexing Program (C&I) and the Federal Depository Library Program (FDLP). Both of these programs became integral to GPO's mission of *Keeping America Informed*.

More recently, with the increase in digital communication and expanding publishing technologies, GPO has streamlined and transformed from a print-only operation to an integrated publishing organization. Recognizing this, Congress re-designated the agency the U.S. Government Publishing Office on December 17, 2014.

The GPO has continually transformed itself throughout its history by adapting to changing technologies. In the ink-on-paper era this meant moving from hand-set to machine type-setting to digital type setting; from slower manual fed presses to high speed presses; and from hand to automated bookbinding.

New strategic priorities for GPO and a *National Plan for Access to U.S. Government Information* for GPO's public information programs are guiding another transformation for Keeping America Informed in the 21st century and beyond.

Bionote

Cynthia Etkin is the Senior Program Planning Specialist in the Office of the Superintendent of Documents (SOD), U.S. Government Publishing Office (GPO). She brings to this position 21 years of experience at GPO and almost 15 years of experience managing Federal depository and law library operations in academic libraries. Cynthia's job focuses on policy and planning for the public information programs of the SOD directly supporting GPO's mission of "Keeping America Informed" and the public's right to freely access its government's information. Recent efforts include preparing for GPO's trustworthy digital repository audit, developing the National Plan for Access to U.S. Government Information, and implementing an eLearning platform. She continues to work on implementing the National Plan while anxiously following Congressional efforts to modernize the statutory authority of the Depository Library Program. Email: cetkin@gpo.gov


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

When is 'grey' too 'grey'? A case of grey data

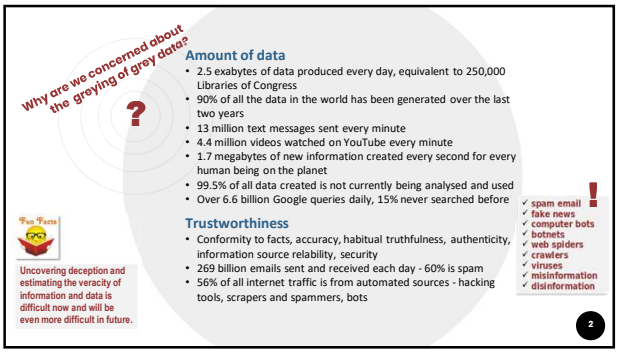
Dobrica Savić, Nuclear Information Section,
International Atomic Energy Agency, NIS-IAEA, Austria

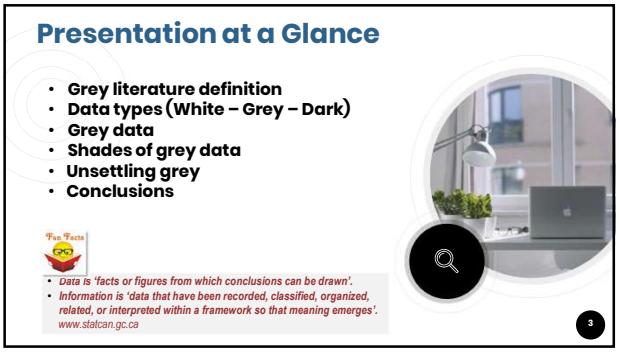
Starting with a definition of grey literature, this paper reviews various document types associated with grey literature with special emphasis on newer types, particularly the data type. Based on a review of relevant literature and current practices, attention is focused on 'grey data', an emerging data type. Its main characteristics, generation, collection, proliferation, importance and use are elaborated. Some moral dilemmas, such as privacy protection and legality, are given extra attention. Suggestions and concluding remarks regarding categorization of 'grey data' as a constituting type of grey literature are also offered.

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 **ORCID iD https://orcid.org/0000-0003-1123-9693**


20th International Conference on Grey Literature
Research Data Fuels and Sustains Grey Literature
3-4 December 2018, New Orleans
WHEN IS 'GREY' TOO 'GREY'?
A case of grey data
Dr. Dobrica Savić
linkedin.com/in/dobricasavic


Why are we concerned about the greying of grey data?
?
Amount of data
• 2.5 exabytes of data produced every day, equivalent to 250,000 Libraries of Congress
• 90% of all the data in the world has been generated over the last two years
• 13 million text messages sent every minute
• 4.4 million videos watched on YouTube every minute
• 1.7 megabytes of new information created every second for every human being on the planet
• 99.5% of all data created is not currently being analysed and used
• Over 6.6 billion Google queries daily, 15% never searched before
Trustworthiness
• Conformity to facts, accuracy, habitual truthfulness, authenticity, information source relability, security
• 269 billion emails sent and received each day - 60% is spam
• 56% of all internet traffic is from automated sources - hacking tools, scrapers and spammers, bots
!
✓ spam email
✓ fake news
✓ computer bots
✓ botnets
✓ web spiders
✓ crawlers
✓ viruses
✓ misinformation
✓ disinformation
Fun Facts
Uncovering deception and estimating the veracity of information and data is difficult now and will be even more difficult in future.
2


Presentation at a Glance
• Grey literature definition
• Data types (White – Grey – Dark)
• Grey data
• Shades of grey data
• Unsettling grey
• Conclusions
Fun Facts
• Data is 'facts or figures from which conclusions can be drawn'.
• Information is 'data that have been recorded, classified, organized, related, or interpreted within a framework so that meaning emerges'.
www.statcan.gc.ca
3

Definition: Grey literature

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

Fun Facts

The first recorded use of grey as a color name in the English language was in AD 700.

4

Data types: White (open) data - Grey data - Dark data

Grey - between the extremes of black and white

Types of grey literature

Bibliographies	Rejected manuscripts	Publications from NGOs and consulting firms
Discussion papers	Un-submitted manuscripts	Videos
Newsletters	Conference abstracts	Wiki articles
PowerPoint presentations	Book chapters	Emails
Program evaluation reports	Personal correspondence	Blogs and social media
Technical notes	Newsletters	Data sets
Publications from governmental agencies	Informal communications	Committee reports
Reports to funding agencies	Census data	Working papers
Unpublished reports	Pre-prints	Company reports
Dissertations	Standards	Catalogues
Policy documents	Patents	Speeches
	Webinars	Reports on websites

- **New sources** of data and information: the Internet of Things (IoT), Machine to Machine communication (M2M), self-driven cars, robots, sensors, security systems, surveillance cameras, and many other systems or apps using AI.
- **Estimated number of currently connected devices** creating specific data vary by billions.
- **Highly contextual and software dependent data and information** is hard to collect and process, and even harder to make sense of and preserve for future use.

Fun Facts

The GreyNet website lists over 150 document types including databases, data sets, data sheets, data papers, satellite data, product data.

5

White (open) data

- Wikipedia
- The International Open Data Charter
- The European Union
- The US Federal Government
- Russia
- China
- Japan

Fun Facts

Open means anyone can freely access, use, modify, and share for any purpose, subject, at most, to requirements that preserve provenance and openness.

6

Dark data

The information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes (e.g. analytics, business relationships and direct monetizing).

Similar to dark matter in physics, dark data often comprises most organizations' universe of information assets. Thus, organizations often retain dark data for compliance purposes only.

Storing and securing data typically incurs more expense (and sometimes greater risk) than value.

By 2020 10% of organizations will have a business unit for making their data commercially available.
(**Gartner**)

Fun Facts

Data mining: the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends.

Data archaeology: preserves historical data using methods for recovering information stored in formats that are becoming (or have become) obsolete.

7

Grey data (GD)

- GD is a type of grey literature that maintains its basic facets such as being recorded, referable, sustainable, valuable, publicly available, and without traditional peer-review.
- GD is an umbrella term that describes the vast array of data that organizations collect and use.
- Useful and valuable data not vetted by peer review, or other governance mechanism.
- GD is often critical to organization's ability to innovate, enhance, and execute its core mission.
- Collected for mandatory or compliance purposes, such as HR, budgets and finance, contracts, procurement, facility management, library users and collections.
- Important for operational, internal management, and legal purposes.
- Data on users, products and services collected for production or marketing purposes.

Facet	White	Grey	Dark
Recorded	x	x	x
Valuable	x	x	x
Referable	x	x	x
Sustainable	x	x	
Used	x	x	
Public	x	x	
Peer reviewed	x		

Fun Facts

Grey or gray is an intermediate color between black and white. It is a neutral or achromatic color, meaning literally that it is a color 'without color'. (Wikipedia)

8

Shades of grey

- **Data in the context of Grey System Theory**
- **Anonymous data as defined by EU**
- **Unstructured data**
- **Unmanaged (risky) data**

Fun Facts

The human eye can only distinguish 32 shades of grey. Computer can display 2^{10} or **1024 shades of grey**. (Robert Fosbinder & Denise Orth, 2012)

9

Unsettling grey

Warning signs!

Data
Unverifiable
Inaccurate (fake) data
Unclear structure
Difficult analysis (format, tools)
Encryption
Redundancy

Purpose
Questionable source
Misinformation
Hidden intent
Data abuse
Defaming

Findable
Accessible
Interoperable
Reusable

Fun Facts

The phrase "shades of grey" usually refers to a situation that is not clear, particularly with regard to whether or not something is categorically evil. When doubt comes into play, things are neither black, nor white, but are in a grey area. (Martha Soren, 2015)

10

Conclusions

- Increased amount of grey data created will impact the way we process, disseminate, manage, and use it
- Increased number of grey data types will demand higher trustworthiness
- Processing needs to be well-thought and present from the beginning of GL data creation. No ad-hoc or post-processing will be possible
- Environmental and technical; economic and financial; as well as social or organizational constraints need to be taken into consideration if long-term grey data sustainability is to be provided
- Usability of grey data depends on the existence of adequate IT tools, the availability of qualified human resources, the protection of intellectual property and the protection of personal privacy
- To secure the future use and maintain the value of grey data, intensive training, wide cooperation, and proper management will be needed
- Only a small percent of businesses extract full value from the data they hold. Use of new IT tools such as AI, might help get more value out of it, improve business results, bring measurable efficiency gains, increase quality of products and services

Fun Facts

The real purpose of data is to uncover patterns, recognize correlations, and identify opportunities that translate to more efficient operations, smarter business decisions, and greater client satisfaction - ultimately, leading to higher profit margins.

11

It's all grey until you find it!
Dean Giustini

Thank you!

12

Big data, grey literature and personal data protection. A crucial impact on social life

Luisa De Biagi, CNR Central Library, National Research Council, Italy;
Roberto Puccinelli, CNR Network and Informative Systems Office,
Antonio Tintori, Institute for Research on Population and Social Policies, IRPPS-CNR, Italy

The European GDPR (EU General Data Protection Regulation) is operative since May 2018. The new EU regulation talks about data regarding identified or identifiable persons owned by Public administration for its institutional missions.
These kinds of data will have to be treated within the board functions, which is obliged to protect them. So, we're not talking about a simple question of 'right to privacy' (the English term is misleading in Italian), but of a clerical obligation going beyond the simple protection of a citizen right.
By consequence, Big Data technologies can be leveraged in order to extract knowledge from grey literature (institutional socials and blogs, patents, health clinical trials, internal documents etc...) but at the same time, they pose serious issues regarding personal data protection.
So, Grey literature assets may contain personal information that big data technologies can extract from huge document bases, thus allowing to acquire knowledge about people's life in unpredictable way.
Our paper aims to analyze the problem, and provide, where possible, useful indications, also with the support of statistical-social research insights.

Bionotes

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 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

Roberto Puccinelli is currently head of Section I at CNR's "Information System Office" and he's been working for CNR since 2001. He has previously worked in the private sector as system and network engineer. As adjunct professor, he has held courses for the First University of Rome "La Sapienza" ("Operating Systems II") and for the Third University of Rome ("Programming and Computing Laboratory"). He graduated in Electronic Engineering at the University of Rome "La Sapienza" and holds a master cum laude in Enterprise Engineering from the University of Rome "Tor Vergata". In the past he has worked in several research projects in the field of Grid technologies both at the national and international level (executive manager of Work Package 11 within the DataGrid project – V Framework Programme, et al.). He's currently involved in the design and development of CNR's information system. In particular, he coordinates projects for the development of application systems and is responsible for the design and implementation of CNR's data warehouse. He is also responsible for CNR's Local Registration Authority management. He's currently involved in projects regarding the design and development of research product open archives and persistent identifier registers/resolvers. He is author of several articles in the fields of Grid technologies, Autonomic Computing, Software Engineering, Open Archives and Persistent Identifiers. Email: roberto.puccinelli@cnr.it

Bionotes *(CONTINUED)*

Prof. Antonio Tintori is a sociologist, Ph.D in Economic geography, CNR researcher and teacher of Methodology of Social Sciences at Sapienza, University of Rome. He was author of several scientific publications and books. His key competences include participation in national and international research projects, analysis of attitudes and behaviors of the young people in cultural and socio-economic fields, such as social integration, deviance, and bullying, well-being, lifestyles and physical activity, development of scenario planning methods and social policies, and competence in scientific disclosure strategies. He was consultant for the United Nations and CNR scientific director of the European Researchers' Night in Rome from 2008 to 2014. His most recent book, with Rossella Palomba, is 'Turn on the light on science. A research-based guide to break down popular stereotypes about science and scientists (https://doi.org/10.5334/bba) (Ubiquity Press, London, 2017). antonio.tintori@irpps.cnr.it

National Research Council of Italy

Big data, grey literature and personal data protection. A crucial impact on social life

by Luisa De Biagi*, Roberto Puccinelli** and Antonio Tintori ***

* CNR-National Research Council of Italy Central Library
** CNR-Telecommunications and Informative Systems Office
*** CNR-Institute for Research on Population and Social Policies (IRPPS)

Grey literature and Big Data

- Grey literature management in large and distributed organizations can leverage Big Data technologies for:
  - acquisition
  - storage
  - Processing and knowledge extraction

Digital preservation systems

- In several countries Public Administrations and private companies must guarantee Long Term Preservation for many different types of Digital Objects
- This can be done in house or acquiring ad hoc services from external providers.

Grey literature and personal data

Grey literature includes many types of documents, structured and unstructured

A Grey literature management system can potentially store all kinds of personal data

Data circulation: pros and cons

Data circulation fuels economy by stimulating research and breeding new ideas.

BUT

can cause harm to individuals and organizations when personal (or confidential) data are used for illegal purposes

General Data Protection Regulation

- In 2016 EU issued the General Data Protection Regulation (EU regulation n. 2016/679)
- The goal is to favour data circulation by providing a solid legal framework which protects individuals from malicious usage of personal data
- It aims at harmonizing national laws in EU countries and defining rules for data export in non EU counties

Principles

- Consent acquisition
- Accountability of the controller
- Data Protection officer
- Risk based approach
- Privacy by design
- Automatic data processing limits

Organisational measures

- An optimization of cybersecurity must be completed by a good privacy policy and data protection: awareness about which data must be storage and their level of truthness in order not to violate people's privacy
- A good SI (Security Intelligence) management and a good organization are equally to be pursued
- Dissemination, divulgation and constant updating about European normatives, rules and informatives, which have to be synchronized together

Organizational measures: the main actors and roles

- Data Controller: Holder of Data Treatment (All EU members)
- Data Processor: Responsible of data treatment
- DPO: Data Protection Officer (**EU Regulation 2016/679). Art.39**. Mandatory for Public Administrations and firms with large scale of sensitive data treatment or involving **large scale**, regular and systematic monitoring of individuals (es. an hospital processing large sets of sensitive data);
- Systems Administrators: though not cited by EU Regulations are important figures to guard security risk and demonstrate responsibility degree

Organizational/Technical measures

- Register of data breaches according to Data breach response and notification procedure
- Creation and updating of cybersecurity breach procedures for immediate notification and in order to determine the entity of the damage and the individuals affected;
- Establishing parameters for risk evaluation and people responsible for making decisions

GDPR: main Social impact and effects

- Personal data protection and privacy of EU citizens as a fundamental human right
- **Right to delete any personal data,** even if it has not public interest, is now always possible
- It's important to determine which are sensitive data (a broader number of cases including, for example, the association to a trade union reported in a pay stub)
- → **focus on social media, cyberbullying, Google and personal accounts**
- breaches impacting on the levels of trust between organisations and customers (organisations breaching citizens' trust will find an increasing commercial disadvantage)

References

- https://www.cybersecurityosservatorio.it/en/Services/statistics.jsp
- http://www.cybersecurityframework.it/sites/default/files/CSR2015_web.pdf
- https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=uriserv:OJ.L_.2018.127.01.0003.01.ITA&toc=OJ:L:2018:127:TOC
- http://ec.europa.eu/justice/data-protection/article-29/index_en.htm
- http://eur-lex.europa.eu/legal-content/CS/TXT/?uri=CELEX%3A32016R0679
- https://www.garanteprivacy.it/regolamentoue

Legal Issues Surrounding the Collection, Use and Access to Grey Data in the University Setting; How Data Policies Reflect the Political Will of Organizations

Tomas A. Lipinski, School of Information Studies; University of Wisconsin-Milwaukee
Kathrine A. Henderson; LibSource, A LAC-Group Company, United States

Universities produce large amounts of grey literature. Grey literature is defined as works that are of sufficient importance to be collected and preserved by the library or its affiliated institutional repository. These works are disseminated through channels other than commercial publishing. While many of these works are protected by intellectual property, intellectual property may not offer a vehicle for protection of grey data. The collector/researcher or his/her home institution may nonetheless consider the data valuable and proprietary (Schöpfel & Lipinski, 2012). Grey literature, and particularly grey data, created at universities gives rise to a number of legal and ethical considerations often addressed through university policy.

Part I of this study describes the legal issues surrounding collection, use, and access to grey data. The authors identify intellectual property schemes, including trade secrets or other proprietary doctrines that may apply. Part 1 also addresses ownership and access rights to data including contract and statutory or regulatory schemes requiring access, or publication of federally funded research in the United States.

Part II examines the grey/open data polices set by institutional repositories Specifically, the authors analyze the Terms of Use/Terms of Access of 11 institutions in the United States and the United Kingdom as reflected in Park and Wolfram (2017) and Park (2018). The analytical framework employed follows Lipinski and Copeland (2015¬) and Lipinski and Kritikos (2017).

Part III considers the tension that results from the need for universities to raise revenue and the public mission/role of the university in society in the same manner as Rooksby (2016.)

Part IV illustrates how data policies reflect the political will of institutions regarding collection, use, and access to open data using the framework found in Post, Raile and Raile (2010). Finally, best practices for future retention, use and dissemination of grey data are forwarded. Consideration is given to collection optimization and access, as well as the minimization of legal risk and uncertainty.

Bionotes

Tomas Lipinski is the Dean of the School of Information Studies at the University of Wisconsin Milwaukee. He 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, Illinois, 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. In 2006 he was the first named Global Law Fellow, Faculty of Law, Catholic University of Leuven, Belgium where continues to lecture annually at its Centre for IT & IP Law and 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. He is active in copyright education and policy-making, chairing PLA Legal Issues in Public Libraries Forum, Chair of the ALA Committee on Legislation Copyright Subcommittee, Chair of the ACRL Copyright Discussion Group from 2013 to 2016, Former Chair and at present member of the ALA OITP Copyright Education Subcommittee, a former member and now Expert Advisor to the Copyright and Other Legal Matters Committee of IFLA and serves as head of an NGO delegation with Permanent Observer status to the WIPO, World Intellectual Property Organization and its Standing Committee on Copyright and Other Rights. tlipinsk@uwm.edu

Bionotes (Continued)

Kathrine Andrews Henderson is research analyst with LAC Group. She is part of a unique team of "virtual" researchers who provide "Library as a Service" to major law firms and corporations. Prior to this Ms. Henderson was the research librarian for the Office of the Auditor General for the State of Arizona. Earlier in her library career, Henderson was as an academic librarian. She was the Instructional Programs Librarian at Thunderbird School of Global Management and served in other roles including time as the business librarian for Arizona State University's Fletcher Library. Kathrine has expertise in business and legal research, intellectual property, and information ethics and has used this expertise to contribute to her field. Recently, she published a chapter on Intellectual Property Ethics in Foundations of Information Ethics, John Burgess and Emily Knox, editors. Other works include co-authoring Case Studies in Library and Information Science Ethics with Elizabeth Buchanan. In January 2018, Henderson was appointed to the Information Outlook Advisory Council for the Special Library Association. In the past, she served as Co-Director of the International Society for Ethics and Information Technology (INSEIT) and as an editor for ACM's Computers & Society. Henderson holds a Masters Degree in Library and Information Science from the University of Wisconsin-Milwaukee and a Bachelor of Science in Management from Arizona State University. In 2017, The School of Information Studies at UWM honored Kathrine as one of 50 Distinguished Alumni as part of the school's 50th Anniversary celebration.

Legal Issues Surrounding the Collection, Use and Access to Grey Data in the University Setting

How Data Policies Reflect the Political Will of Organization

Tomas A. Lipinski
Kathrine A. Henderson

Introduction

- Grey data are works of sufficient importance to be collected and preserved by the library or its affiliated institutional repository.
- Several university policies were studied asking whether or not these policies expressed or reflected the "political will" of the institution towards its grey data.
- The study discusses: political will, legal issues, analysis of data policies, conflicting roles and recommendations.

Defining Political Will

Post, Raille, & Raille (2010) developed a systematic approach defines "political will" into successive components:

1. A sufficient set of decision makers
2. With a common understanding of a particular problem on the formal agenda
3. Is committed to supporting
4. A commonly perceived, potentially effective policy solution

Legal issues surrounding collection, use, and access to grey data -- Copyright

- Facts: survey or experiment results, statistics, field observations / recordings
- Compilations of data
- Works of thin copyright:
  - Oral histories and interviews
  - Mixed works: lab results or field observations with researcher commentary / notes
- Fair Use:
  - Re-use of "thin works" and
  - Reproduction in order to extract unprotected data or data-mining

Legal issues surrounding collection, use, and access to grey data – Trade Secrets

- Elements:
  - Independent economic value, actual or potential
  - Value derives from not being generally known
  - Not be readily known or ascertainable (a secret)
  - Reasonable measures must be taken to secure the content
  - Not ascertainable by proper means
    - Proper means: reverse engineering
    - Improper means: espionage, theft, bribery, misrepresentation and breach or inducement of a breach of duty to maintain secrecy. Wis. Stat. § 134.90(1)(a)
  - The information must have economic value to the party appropriating the secret
- Would not apply to grey data generated from federally (U.S.) funded projects

Legal issues surrounding collection, use, and access to grey data – Misappropriation

- *International News Service v. Associated Press*, 248 U.S. 215 (1918)
- *National Basketball Association v. Motorola, Inc.*, 105 F.3d 841 (2nd Cir. 1997)
  - Information gathered at a cost
  - Time-sensitive
  - Defendant's use of the information constitutes free-riding
  - In direct competition with the plaintiff
  - Free-riding reduces the incentive to produce the product or service

US Universities Data Policies & Terms and Conditions: Analysis

- Subject of Terms and Conditions: IL (copyright focus), MI, Harvard, MN and Purdue (most elaborate) govern users, VA Tech (contributors)
- Contract formation:
  - Click-to-agree (MI)
  - Use equals assent (IL, MN, VA Tech, Purdue)
  - No contract (Harvard)
- Attribution (moral right): attribution (all except Purdue)
  - CC BY or statement suggesting that result

US Universities Data Policies & Terms and Conditions: Analysis

- Permissions (all except Purdue): Illinois and Harvard (CCO or non-standard), Michigan (akin to CC SA), Minnesota (instructs users to contact authors) and VA Tech (seek permission of author to upload)
- Privacy: Michigan, Minnesota, VA Tech (IRB protocols / Administration responsible for user privacy) and Purdue
  - Illinois and Harvard: no indication
  - Tracking technologies: Illinois and Harvard (Google Analytics) and Purdue (cookies)
- Use restrictions: MI (NACJD and NCHS), Harvard (no contract, CCO alone), Purdue (elaborate), IL and MN (none)

US Universities Data Policies & Terms and Conditions: Analysis

- Disclaimers (all but Harvard), MI (data collector and funding agency responsible), MN (warranty or fitness for a particular purpose), VA Tech (provides warranty?: contact Administrators issues/errors), Purdue (“as is”, fitness, merchantability, accuracy, completeness, uninterrupted, errors, viruses or harmful components)
- Damage waivers: MN and Purdue
- Sanctions: MI (return data, deny access, tenure/termination, institutional suspension/damages), VA Tech (right to remove), Purdue (termination/suspension)

Conflicting Roles

- Protecting grey data is subject to the same sorts of difficulties that arise whenever higher education seeks to protect its intellectual outputs.
- Data policies should, and in some circumstances, must require that the intellectual benefits inherent to grey data flow into society; however, there is a competing narrative.
- Exercising intellectual property rights is lucrative and helps to ensure institutional longevity; however, it also requires institutions to defend their property rights.

Conclusions & Recommendations

- Employ click-to-agree for terms and conditions of use
- Disclaim warranties regarding the reliability of the data
- Exercise relevant / applicable proprietary rights:
  - Use CCO unless copyrightable content or compilation data set
  - Use CC BY license or a provision requiring proper citation/attribution
- Address other legal obligations:
  - Privacy of research subjects
  - Publication of publicly funded research, and
  - Contractual agreements with private partners.
- Should reflect the educational mission of the university

Questions???

THANK YOU

Tomas A Lipinski
and
Kathrine A Henderson

Creating, using, and distributing compliant research data: A model for fulfilling legal and policy requirements

Daniel C. Mack, University of Maryland, United States

Research data occur in a wide variety of formats and media and cover all disciplinary areas. Some types or classes of data may have specific requirements that vary with content or purpose. However, most data need to comply with some set of policies, laws, or other guidelines. Individual institutions may have their own policies governing creation, use, and management of data. Likewise, grantors, sponsors, and other stakeholders may also require research data to follow their own guidelines. In addition, local, national, and international law may have implications for the use of research data. All of these issues can be complicated when researchers collaborate among multiple institutions, often in a variety of legal jurisdictions.

When creating, using, and disseminating data, researchers must comply the requirements of various types of data, legal and policy compliance, jurisdictions, and other factors. Fortunately, resources are available to guide researchers through this often complicated landscape of compliance. By identifying the proper tools, personnel, and other resources from the beginning of a research project, investigators can ensure that they create and manage data in ways that comply with all necessary requirements. This paper presents a model for navigating the often complex problems posed by the necessities of dealing with multiple data types, legal and policy requirements, overlapping jurisdictions, and data uses. It will also identify a suite of freely available tools to assist researchers in the creation, use, and distribution of research data that meets the compliance needs of individual research projects.

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 **ORCID iD https://orcid.org/0000-0002-4351-6897**

UNIVERSITY LIBRARIES

Creating, using, and distributing compliant research data: a model for fulfilling legal and policy requirements

20th International Conference on Grey Literature
New Orleans, Louisiana, USA
3 December 2018

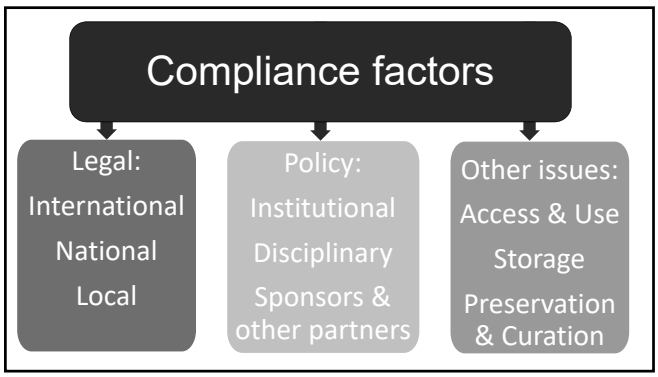
GreyNet

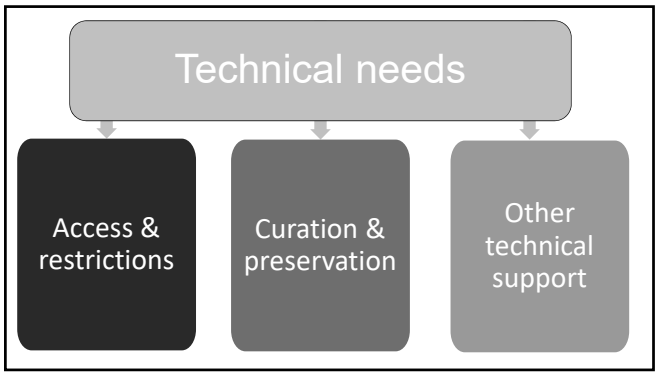
Creating and documenting compliant data

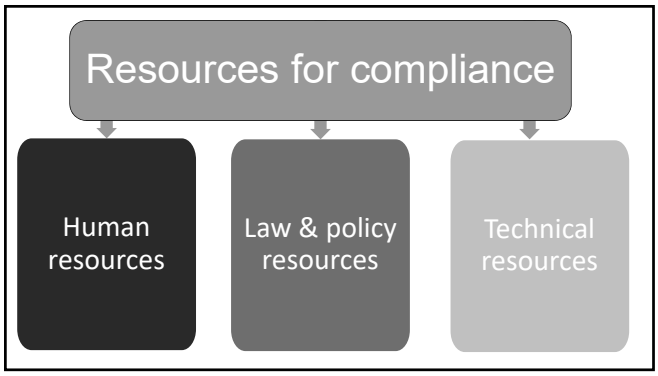
- Identify types of data
- Identify compliance factors
- Identify technical needs
- Identify resources

Types of research data

- Quantitative vs Qualitative
- Restricted vs Unrestricted
- Data content, format and other factors


Compliance factors
Legal:
International
National
Local
Policy:
Institutional
Disciplinary
Sponsors & other partners
Other issues:
Access & Use
Storage
Preservation & Curation


Technical needs
Access & restrictions
Curation & preservation
Other technical support


Resources for compliance
Human resources
Law & policy resources
Technical resources

Restricted and open research data

- Many legal and regulatory requirements for sponsored research involve the extent to which data is restricted or open
- Many government and private sponsors of research require that information, including data, be open and freely available to other researchers
- Much sponsored research also involves data with a wide range of restrictions on access, use, and identification of source
  - HIPAA
  - FERPA

Selected resources: human & technical

- Human resources: researchers, administrators, counsel, technical and clerical support personnel, other stakeholders; varies widely by institution and organization
- University of Maryland Guide to Writing Data Management Plans: https://www.lib.umd.edu/data/dmp-guide
- Research Data Management at Pitt: https://pitt.libguides.com/managedata
- 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."

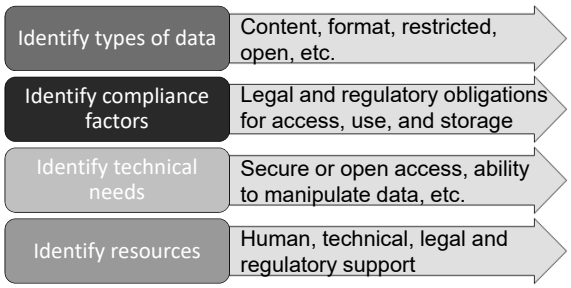
Selected resources: law and policy

- Europe/EU:
  - Law & regulation: https://ec.europa.eu/info/research-and-innovation/law-and-regulations_en
  - EU Open Data Portal: https://data.europa.eu/euodp/en/home
- UK Data Service: https://www.ukdataservice.ac.uk/manage-data/plan/checklist
- International Standards Organization (ISO): iso.org

Selected resources: law and policy continued

- FERPA: https://studentprivacy.ed.gov/?src=fpco
- HIPAA: https://www.hhs.gov/hipaa/index.html
- NSF: https://www.nsf.gov/bfa/dias/policy/
- NIH: https://grants.nih.gov/policy/index.htm
- NEH: https://www.neh.gov/grants/manage
- NEA: https://www.arts.gov/grants/manage-your-award
- SPARC directory of federal funder sharing policies: http://datasharing.sparcopen.org/articles

Model: create & document compliant data



Compliant research data

- Fulfills statutory and regulatory obligations of institutions, organizations, sponsors, and governments
- Maximizes impact on scholarship because it meets requirements for access, distribution, and re-use
- Documents due diligence on part of researchers

Questions or comments? Daniel Mack: dmack@umd.edu

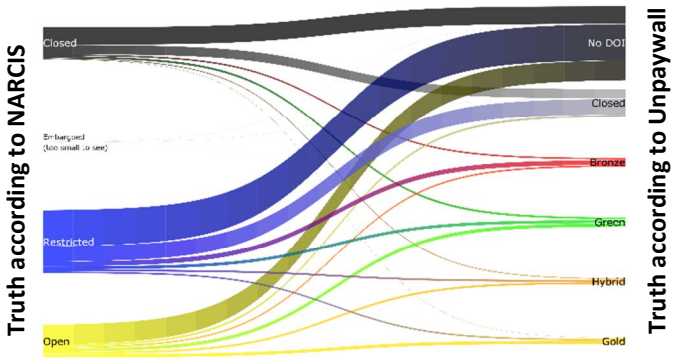
On Open Access to Research Data: A DANS Use Case

Emilie Kraaikamp, Chris Baars, Henk Harmsen, and Hella Hollander
Data Archiving and Networked Services, DANS-KNAW, Netherlands

The Dutch government propagates Open Access for all (research) publications and data that has been funded with public money. There is an official target, that in 2024 all publicly funded data should be Open Access available. At this moment in the Dutch scientific landscape, we see that this is for publications between 40-60% and for data 10-50%.

With these numbers in mind, while digging deeper in the definition, you may ask yourself 'what is Open Access?' Comparing the information of local CRIS systems in the Netherlands (published by the DANS service NARCIS) with the information of Unpaywall, learns that Open Access in the one system is not always Open Access in the other system. Cynics call these access categories Trial and Error Access.

Chris Baars/Emil Bode cc0-2018



Another example. Amongst the usual standard licenses, DANS developed for its own archive a specific Open Access license (Open Access for registered users, OAFRU), which is not recognized as Open Access by Europeana during their harvesting process of our archive.

In this paper I would like to zoom in on the definitions of Open Access and compare this to the openness of use. Something can be open in its access, but at the same time it can be restricted in its use. Or the other way around: something can be restricted in its access, but open in its use. Restricted access is always behind some kind of wall; be it a pay-wall or a login-wall. Access right are about accessibility, licenses are about the (re-)use. I will give some special attention to the embargo option, as being a possible Open Access category.
The next step then is comparing the 'hard-core' Open Access license CC0 with its other CC-relatives.
Then I will focus on the differences between Open Access for publications and Open Access for data. And last but not least, I will give some recommendations for the use of Open Access licenses for Research Data.
To clarify my arguments, I will use examples from the DANS Archive.

Data Archiving and Networked Services
DANS

Open access and research data

A DANS use case

Emilie Kraaikamp, adviser legal affairs

December 3, 2018. GL20 Conference, Loyola University New Orleans, Louisiana USA

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

Overview

- Introduction to DANS
- Open access: definitions
- Publications and data: differences in open access
- Creative commons CC0 and licences: what do they offer
- Choosing open access for research data: a process
- Recommendations for researchers
- Conclusions

Data Archiving and Networked Services
DANS

Introduction to DANS

DANS (Data Archiving and Networked Services) is the Netherlands Institute for permanent access to digital research resources. DANS encourages researchers to make their digital research data and related outputs Findable, Accessible, Interoperable and Reusable.

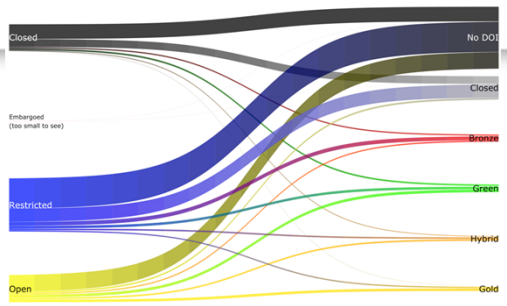
We provide expert advice and certified services. Our core services are: DataverseNL for short-term data management, EASY for long-term archiving, and NARCIS, the national portal for research information.

By participating in (inter)national projects, networks and research, DANS contributes to continued innovation of the global scientific data infrastructure. Open if possible, protected where necessary. DANS is an institute of the Dutch Academy KNAW and funding organisation NWO.

Data Archiving and Networked Services
DANS

Open access: definitions

- The open access movement
  - Markers: Budapest Open Access Initiative, the Bethesda Statement on Open Access Publishing, and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
- Initially focus on publications
- Open access: refers to both access and reuse



Open access, interpretations 2018. On the left: truth according to local CRIS Systems in the Netherlands (via NARCIS). On the right: truth according to Unpaywall. *Image: CC0, credits: Emil Bode (DANS).*

Publications and data: differences in open access

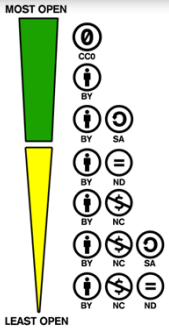
- Publications: a single object
- Research data: an often diverse collection
- Open access for research data means providing access to a great range of material and therefore a careful consideration for reuse

Creative commons CC0 and licences

The CC0 tool: CC0

The six licences:

- Attribution CC-BY
- Attribution ShareAlike CC BY-SA
- Attribution-NoDerivs CC BY-ND
- Attribution-NonCommercial CC BY-NC
- Attribution-NonCommercial-ShareAlike CC BY-NC-SA
- Attribution-NonCommercial-NoDerivs CC BY-NC-ND



DANS

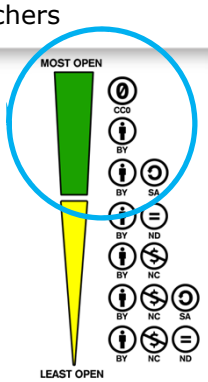
Choosing open access for research data

- Background: the open access movement and national policy.
- Barriers and understanding the benefits
- Open access at DANS, implementation, experiences.

DANS

Recommendations for researchers

- CC0 is best fit for research data
- DANS promotes CC0 and offers as direct alternatives CC-BY and CC-SA: the three less restrictive options for reuse.



DANS

Conclusions

- Open access relates to both access and reuse.
- Opposed to publications, research data can be diverse and extensive and require a more careful approach in applying open access.
- Creative commons offers suitable means to make research data open access
- Positive developments at DANS regarding research data and open access
- CC0 is best to use for research data, with CC-BY and CC-SA as good alternatives

Data Archiving and Networked Services
DANS

Data Archiving and Networked Services
DANS

Thank you for your attention.
Questions?

emilie.kraaikamp@dans.knaw.nl
Twitter: @dans.knaw.nl
dans.knaw.nl/en

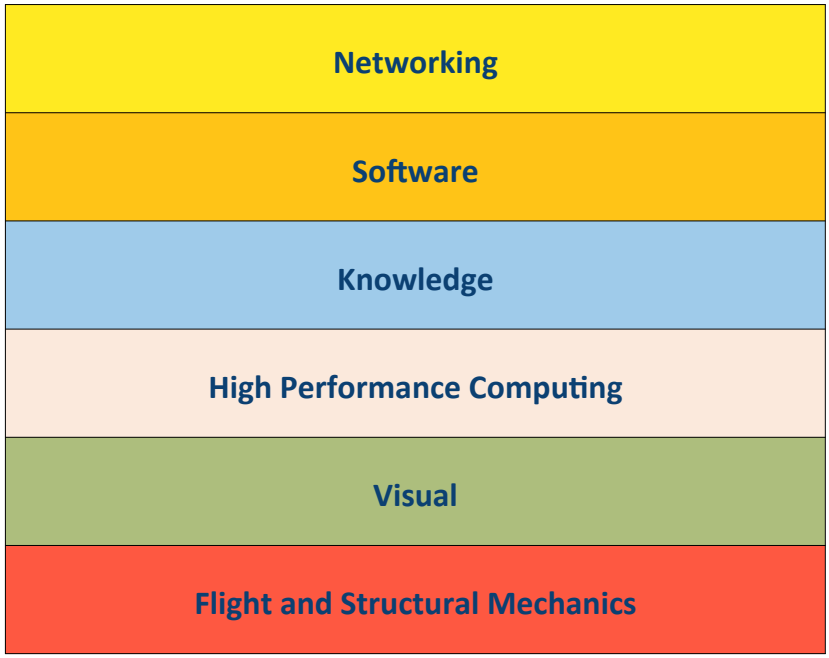
dans.knaw.nl
DANS is an institute of KNAW en NWO

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 Richerca del CNR

Contact: +39 050 315 2403
segreteriascientifica@isti.cnr.it
http://www.isti.cnr.it

The data librarian: myth, reality or utopia?

Silvia Giannini and Anna Molino,
Institute of Information Science and Technologies, ISTI-CNR, Italy

The emergence of e-science and e-research has opened new paths and trends in scholarly communication and management. In the academic environment, the need for opening research products to a wider audience has become increasingly urgent. In this perspective, the Open Science (OS) movement is growing considerably in academia and among scientists worldwide. Two fundamental aspects of OS are the Open Access (OA) to scientific publication and the possibility of discovery, sharing and exploit the data used for or produced during the research process. The need for creating Open Data (OD) is profoundly changing the perspectives adopted by researchers during the scientific production, as research data is increasingly recognized as a primary research output.

Not surprisingly, in order to have access to grants, funding bodies such as the European Commission, the Welcome Trust and the RCUK in UK, the Australian Research Council in Australia, or the National Institutes of Health in the U.S. ask for a Data Management Plan (DMP) accompanying the project proposal[1].

In this perspective, many academic libraries are now extending their century-long track record in the professional management of knowledge resources towards the area of research data, seeking to maximize research data skills among staff in their organisations[2]. Academic libraries are indeed increasingly involved in the management of research data across the lifecycle[3], actively participating in tasks such as providing access to data, supporting researchers in managing their data and drafting DMPs, as well as managing data collections. Given this framework, it becomes clear that the "librarian" represents a constantly evolving profile.

Our work will propose an overview of the competencies required to the librarian operating in academic libraries nowadays, delivering examples and making comparisons between different experiences worldwide. Special attention will be granted to the educational aspects necessary to accomplish with the actual skills required to research librarians, being training a fundamental aspect in the development and definition of this profession, highlighting the similarities and differences in educational proposals among different academic institutions.

We will focus on the following questions: what is currently missing in the background of librarians? Should it be implemented or we must consider the *data librarian* as a brand-new profile? Our aim will be trying to give possible answers and outline specific qualifications required to those currently operating in academic libraries.

References

[1] Cassella M. (2016). *Dal digital curator al data librarian. Biblioteche oggi – Speciale Digital Update*, 34 (4) 13-21.

[2] Research Data Alliance – Libraries for Research Data IG (2015). *How to maximize research data skills in libraries*, https://www.rd-alliance.org/how-maximize-research-data-skills-libraries.html.

[3] Schmidt B., Shearer K. (2016). *Librarians' Competencies Profile for Research Data Management.* COAR Joint Task Force on Librarians' Competencies in Support of E-Research and Scholarly Communication, https://www.coar-repositories.org/activities/support-and-training/task-force-competencies/

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
ORCID iD https://orcid.org/0000-0001-7323-3786

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.

20th International Conference on Grey Literature - December, 3-4 2018
Loyola University, New Orleans, Louisiana USA

The data librarian: myth, reality or utopia?

Silvia Giannini, Anna Molino
CNR-ISTI, Italy

Outline

1. Scenario and objectives
2. What is research data?
3. The Research Data Management (RDM)
4. Digital curation
5. Which professional figures?
6. Case studies
7. The *data librarian* profile
8. Conclusions

Scenario

- Open Science
  - Open Access to scientific publication.
  - Open Data as a *primary research output*.
- Requests for services and infrastructure to manage and preserve research data from many funding bodies .
- Academic libraries as "*aggregator, collector and curator of external scholarship, be it printed or online*".

Objectives

- Who and how is managing the research data?
- What experiences worldwide?
- What roles for academic libraries?

What is research data?

- The definition and the management of research data are depending on the differences between disciplinary field and organizational context
  - *observations, images, computer program results, recordings, measurements, experiences, theory, test or hypothesis, or another research output is based…*
- Research data comes in a variety of formats
  - *numerical, descriptive, visual or tactile…*
- It may be raw, cleaned or processed.
- It may be held in any format or media.

Research data is the material underpinning a research assertion

The Research Data Management (RDM)

<<Research data management concerns the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information.>> Whyte A., Tedds J. (2011)

RDM as Research Data Service (RDS)

- Data Management Plan
- Digital curation
  - *selection*
  - *preservation*
  - *maintenance*
  - *archiving*
- Metadata creation and transformation

Research Data Management Lifecycle, diagram
The University of California

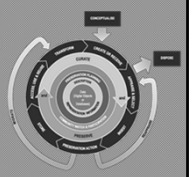
Digital curation

- The concept considers the ubiquity of digital contents and their need to be preserved, managed and valued.
- *Digital curation* is
  - a broader concept;
  - a *cross-activity* among different domains;
  - involving the entire *lifecycle of a digital resource.*

- *The DCC model summarizes the full lifecycle actions in:*
  - description and representation information;
  - preservation planning;
  - community watch and participation;
  - curate and preserve.

The need to identify a professional figure who manage and store the growing amount of data in digital format has generated the role of *digital curator*.

Digital Curation Centre (DCC)
Curation Lifecycle Model



www.dcc.ac.uk

Which professional figures?

The *digital curators* community

The link of the *digital curation* with the development and management of Open Access repositories makes *data curation* a subset of the *digital curation*.

At the same time, we can consider *data curator* a specialization of the *digital curation* figure, while...

Thus, data curation is
- a narrower concept;
- strongly connected to the academic world;
- referred to data as *research data*;
- involving the entire *lifecycle of digital research data.*

...the wide experience of librarians in different disciplinary domains, their skills in the management of metadata sets and in the selection, care and maintenance of collections directly refers to the figure of the *data librarian*.

Case study – University of Edinburgh Research Data Service (RDS)

RDS: tools, support and training to aid staff and students working with research data

RDM support and consultancy
- Before: Data Management Plan.
- During: data search, storage, sharing, and versioning.
- After: datasets recording; open data sharing; data archiving.

Training
- Online:
  - MANTRA[1] free online course for those who manage digital data.
  - Special focus "DIY Training Kit for Librarians[2]."
  - RDMS MOOC [3](Massive Open Online Course) in collaboration with the University of North Carolina.
- Face to Face: workshops and courses.

[1] https://mantra.edina.ac.uk/ [2] https://mantra.edina.ac.uk/libtraining.html [3] http://www.coursera.org/learn/data-management

Case study – The RDNL FO-BO Model

Research Data Netherlands (RDNL): alliance of DANS, 3TU.Datacentrum and SURFsara.

Services provided: data curation, management and storage.

Front Office (FO)
- Data management: information provision and researchers training.
- Temporary storage facilities and data transfer.
- Data acquisition.

Back Office (BO)
- Data management: data stewardship, long-term storage and accessibility.
- Training and support to FO data librarians.
- Secure sustainable storage and retrieval at the end of the research project.

- Specific FO-BO responsibilities vary from organization to organization.
- Duties of front and back offices may overlap: coordination and role definition are necessary.
- Institutes performing FO tasks only may outsource their BO tasks to a data archive.

Case study – eResearch at Griffith University

Division of Information Services (INS): library and Information Technology (IT) services.

Librarians: acquisitions, collection development, copyright advice, IT training, OA, publication repositories, research assessment exercises, bibliometrics.

eResearch Services: build and manage technical infrastructures for supporting researchers.

Librarians: "generalists" as they need to have a wide range of skills, knowledge and expertise

Core set of skills
- Advanced metadata skills
- High level communication skills
- High level documentation skills

Core set of knowledge
- Knowledge of the broader researcher environment
- Knowledge of the research process and of scholarly communication
- Knowledge of legal and regulatory framework

Generic skills and knowledge
- Technical skills (e.g. system design, business analysis, usability testing)
- Project Management to understand project lifecycles, PM methodologies, internal PM frameworks and reporting requirements

The *data librarian* profile

<<"Traditional" librarian competences and skills into renewed organizational structures.>> (*Schmidt et al., 2016*)

Core competences (Schmidt & Shearer, 2016)
- Provide access to data.
- Advocacy and support for managing data.
- Managing data collections.
- Related services: OA and institutional repositories; collection development; advisory services (e.g. copyright policies); information literacy; digital curation and preservation; digital collections.

eResearch role (Brown et al., 2015)
- Understanding of discipline-based research process, outputs and scholarly communication (e.g. data types and formats).
- Knowledge of ethics, intellectual property, copyright and licensing.
- Knowledge of discipline-specific metadata schemas and related standards (item and collection level).
- Knowledge of repository certification schemes and standards.
- Knowledge of semantic web standards.

Conclusions

Echo and Narcissus by John William Waterhouse

The Radical Christianity of Thomas More's "Utopia"

- The importance and need for data management is recognised and established by national entities and funding bodies.
- It arises from the case studies that having a formal RDM Policy leads to the definition of specific roles involved in RDS.
- While roles and professional skills of the *digital librarians* can be often considered as re-interpretations of traditional roles, the role of the *data librarian* is one of the most complex and should be defined completely.
- The worldwide scenario shows data librarians operating in various contexts with different backgrounds.

So, *neither myth… nor utopia …but a composite reality*

Library, Information Science & Technology Abstracts™ with Full Text

Available via EBSCO*host*®

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

Research Data Management: What can librarians really help?

Yuan Li, Willow Dressel and Denise Hersey
Princeton University, United States

In 2013 the United States Office of Science and Technology Policy released a memorandum directing federal agencies with more than $100 million in research and development funding to develop plans to, among other things, make scientific research data from unclassified research publicly accessible. At the same time 74% of libraries surveyed by the Association of Research Libraries indicated they were already engaged in research data management services. Since the 2013 memo, both government and journal data management requirements have increased and librarians have continued to use their expertise in information organization, management, and preservation to help meet this growing need. This paper will present a case study on data management services developed and delivered by librarians at Princeton University, including providing services for the Department of Energy's Princeton Plasma Physics Laboratory. Service development was started by a single science librarian but eventually grew to include the scholarly communications librarian and relevant science subject librarians. Services include data management plan consultation, data management education, and helping connect researchers to the institutional data repository. Services have been developed and delivered using existing staff. It is expected that dedicated staff will be hired in the future to fully develop the program. Staffing costs vary based on the individual institution and location. The paper will highlight outcomes of selected collaborations, discuss feedback on services, and provide a case study for peer institutions.

Bionotes

Yuan Li - As a Scholarly Communications Librarian at Princeton University, Yuan Li manages the Princeton University Library's efforts to support scholarly publication innovations and reforms, and supervises and coordinates activities related to the Princeton Open Access policy and Repository. Previous roles include Scholarly Communication Librarian at Syracuse University, Digital Initiatives Librarian at University of Rhode Island, and Digital Repository Resident Librarian at University of Massachusetts Amherst. She holds an MLS from the University of Rhode Island, a MS in Applied Computer Science from the National Computer System Engineering Research Institute of China, and a BS in Computer Science from Yanshan University (China). As a Scholarly Communications Librarian at Princeton University, Yuan Li manages the Princeton University Library's efforts to support scholarly publication innovations and reforms, and supervises and coordinates services related to open access, copyright, and data management. Previous roles include Scholarly Communication Librarian at Syracuse University, Digital Initiatives Librarian at University of Rhode Island, and Digital Repository Resident Librarian at University of Massachusetts Amherst. She holds an MLS from the University of Rhode Island, a MS in Applied Computer Science from the National Computer System Engineering Research Institute of China, and a BS in Computer Science from Yanshan University (China).

Denise Hersey holds an MA in American History from University of Massachusetts, a BA from University of Pennsylvania, and an MLS from Southern Connecticut State University. She is currently the Head of the Science Libraries at Princeton University, having previously held the position of Assistant Director of Clinical Information Services at the Cushing/Whitney Medical Library at Yale University. Denise has also worked as a corporate librarian and at liberal arts institutions. Her current professional interests are in the communication of science to the public, and using user experience methods to help inform changes within libraries.

Willow Dressel, Engineering Librarian at Princeton University

Princeton University LIBRARY

Research Data Management
Can Librarians Really Help?

Yuan Li, Scholarly Communications Librarian
Willow Dressel, Engineering Librarian
Denise Hersey, Head of Science Library

Princeton University Library
The 20th International Grey Literature Conference
December 3-4, 2018

PRINCETON UNIVERSITY

Research Data Management Lifecycle (UK Data Archive)

What is data management?

Differences among disciplines

PRINCETON UNIVERSITY

Princeton University

- Undergraduate Enrollment - 5,260
- Graduate Enrollment - 2,845
- Faculty - 1,261
- Student to Faculty Ratio - 5:1
- Areas of Study: Humanities, Natural Sciences, Social Sciences, Engineering and Applied Sciences
- Princeton University Library - 10 branches, over 7 million books

Image Credit: Denise Applewhite
Princeton Office of Communications

PRINCETON UNIVERSITY

Data Management Planning

Librarians can:

- Raise awareness on funder mandates and the benefits of planning
- Assist in creating data management plans or recommend templates
- Educate on good data management practices
- Offer customized support for specific projects

PRINCETON UNIVERSITY

Collecting Data

Librarians can:

- Provide data reference
- Collect, purchase, and license data sets
- Determine best practices for research file organization and naming
- Advise on documentation and data level metadata schema
- Educate or refer on legal or ethical issues with collecting data
- License or support electronic lab notebooks or qualitative tools such as Redcap or Nvivo

PRINCETON UNIVERSITY

Processing and Analysing

Librarians can:

- License data analysis tools
- Offer training on data analysis methods and visualization
- Advise on data documentation and processing metadata
- Provide guidance or support on storage, back-up, and versioning
- Educate on different software formats for preservation

PRINCETON UNIVERSITY

Publishing and Sharing

Libraries can:

- Provide and recommend options for publishing and sharing including relevant repositories
- Host and manage open access repository
- Help researchers understand and comply with funder public access mandates
- Curate data
- Assist with pubic dissemination and marketing of data
- Instruct on data rights management

PRINCETON UNIVERSITY

Curating/Preserving

Libraries can:

- Offer mid-term and long-term preservation strategies
- Advise on/Monitor file formats
- Generate permanent identifiers for data and files
- Run file identity checks

PRINCETON UNIVERSITY

Re-using

Librarians can:

- Provide support in Geographic Information Services and GIS data
- License pre-existing data sets
- Collaborate on departmental data archiving policies
- Instruct on data citation
- Educate or refer on legal or ethical issues with collecting data

PRINCETON UNIVERSITY

Challenges and Opportunities

Challenges:

- Staffing
- Expertise
- Technology

Opportunities:

- Increase library's integration to research
- Development of new courses in MLS programs

Opportunities in Princeton University Library:

- Collaborative mini course for graduate students - January 2018
- Opening for Director of Research Data Services
- Collaboration with other campus partners, such as OIT and Office of Research

PRINCETON UNIVERSITY

Final Message

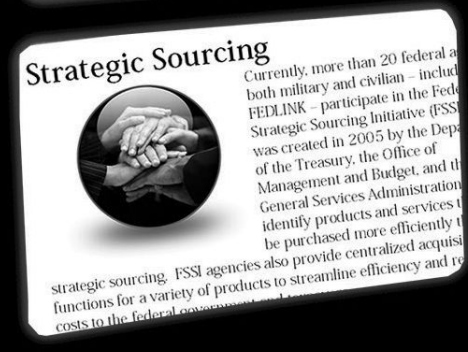
- Libraries are able to provide a spectrum of support
- Resources (personnel and monetary) are needed for providing such support
- But you don't have to wait, you can start with basic services and build up from there
- It's never too late to start Data Management Services

PRINCETON UNIVERSITY

Works Cited

- "Research Data Lifecycle." UK Data Service. Accessed October 30, 2018. https://www.ukdataservice.ac.uk/manage-data/lifecycle.
- National Science Foundation. "Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century." National Science Foundation, 2005. https://www.nsf.gov/pubs/2005/nsb0540/.
- Carleson, Jake. "The Use of Life Cycle Models." In *Research Data Management: Practical Strategies for Information Professionals*. Charleston Insights in Library, Archival, and Information Sciences. Purdue University Press, 2014.
- Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. "Sharing Detailed Research Data Is Associated with Increased Citation Rate." *PloS One* 2, no. 3 (2007): e308.

PRINCETON UNIVERSITY



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.

LIBRARY OF CONGRESS
FEDLINK

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

Data Management and the Role of Librarians

Plato L. Smith and Sara Gonzalez
George A. Smathers Libraries; University of Florida, United States

'Research Data Science' is defined by CODATA-RDA as an ensemble of (1) Open Science principles and practices (FAIR) and research data management and curation skills, (2) the use of a range of data platforms and infrastructures, (3) large scale analysis, (4) statistics, (5) visualization and modeling techniques, (6) software development and annotation, and (7) more. Data management and the role of librarians must now include developing expertise with and training faculty, students, and staff on "research data science" directly and/or indirectly through collaborative library/faculty partnerships. To meet this need, librarians at the University of Florida have developed a new research support service called Academic Research Consulting & Services (ARCS) to assist faculty, students, and staff with their data management and research needs. The library-centered UF Data Management and Curation Working (DMCWG) and ARCS work in collaborative partnerships with the UF Informatics Institute, UF Data Carpentry Club (https://github.com/UF-Carpentry), and UF Data Science & Informatics (DSI) undergraduate student organization to provide support to pre- and post-grant research and teaching.

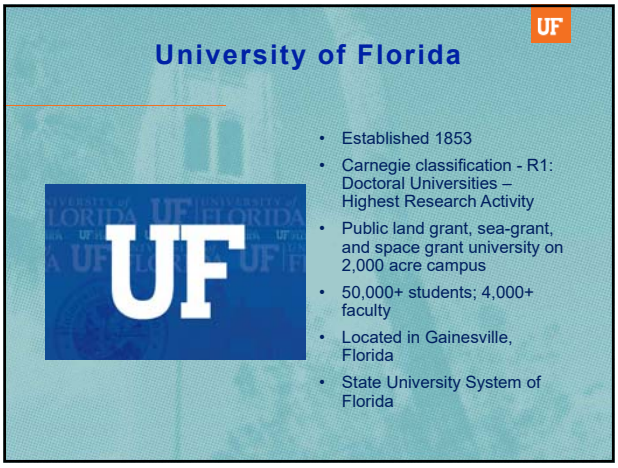
This new role of librarians is to facilitate library/faculty collaborations and broker resources that contribute to the facilitation of promulgating 'research data science' skills at scale for their respective institutions. The paper will discuss the ongoing outreach activities, inter-departmental collaborations, and future goals or leveraging capacity, infrastructure, and resources to develop and teach research data science at scale within an institution's current organizational culture. One of the main goals of this paper is to highlight the importance and significance of developing successful library/faculty collaborative relationships built on character, integrity, and humility as the cornerstones for the roles of librarians in promoting good data management and FAIR principles.

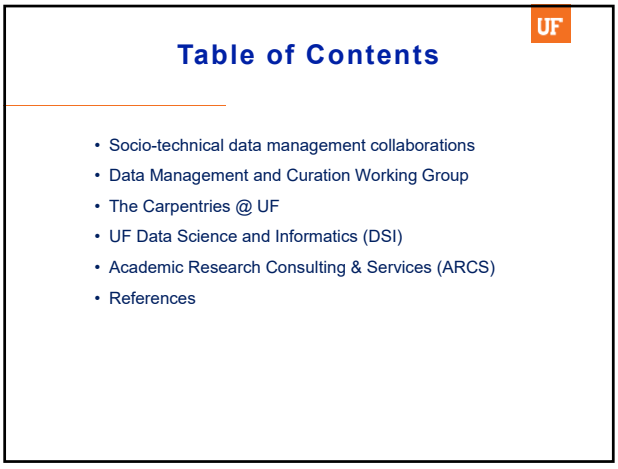
Bionote

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 **ORCID iD https://orcid.org/0000-0003-1814-0151**


UF
Data Management and the Role of Librarians
GL20 – Twentieth International Conference on Grey Literature
Plato Smith and Sara Gonzalez
December 3, 2018
Loyola University
New Orleans, Louisiana USA


UF
University of Florida
UF
Established 1853
Carnegie classification - R1: Doctoral Universities – Highest Research Activity
Public land grant, sea-grant, and space grant university on 2,000 acre campus
50,000+ students; 4,000+ faculty
Located in Gainesville, Florida
State University System of Florida


UF
Table of Contents
Socio-technical data management collaborations
Data Management and Curation Working Group
The Carpentries @ UF
UF Data Science and Informatics (DSI)
Academic Research Consulting & Services (ARCS)
References

Socio-technical data management collaborations

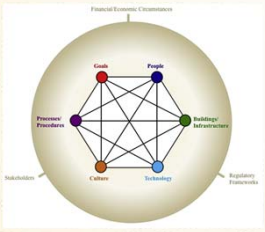


Fig. 1 – Social-technical system framework (Davis et al., 2014)

- UF Marston Science Library
- UF Office of Research
- UF Informatics Institute (UFII)
- UF Research Computing
- UF College of Engineering
- UF Institute of Food and Agricultural Sciences (UF/IFAS)
- UF Clinical and Translational Science Institute (UF/CTSI)
- Data Sharing Policies
  - *NIH data sharing comments*
  - *Cooperative agreements*
  - *PLOS Governing Board sensitive data*

Data Management and Curation Working Group

- 2012 – 2016 Task Force
- 2016 – Present Working Group
- 52 subscribers
- Composed of Chairs, faculty, staff, IT
- Meets every 4th Monday in Marston Science Library
- Contributes to development of campus initiatives (e.g., ORCiD, research grants, training)
- Partners with The Carpentries @ UF, DSI, ARCS, and others

- ✓Several general data management training workshops
- ✓Average attendance 12
- ✓Several discipline-specific training workshops
- ✓DMP support – $6.2 m
- ✓1st Annual Data Symposium (3/19/18)
- ✓Two sub awards (UF/IFAS)
- ✓Lead on UFDMAC (NIEHS)

The Carpentries @ UF
(https://www.uf-carpentries.org/)

Carpentries @ UF

- Established 2016
- Carpentry premium member
- Developed UF Carpentries Club and Board in 2017
- Extended collaborations with UFII, UFBI, UF Marston Science Library, iDigbio, UF Florida Museum of Natural History (FLMNH), Wildlife Ecology, and the White Lab

2018 Workshops and Other Events (total of 18 thru 10/24/2018)

- R for Geospatial Data
- One Day R Workshop
- Research Bazaar
- Data Carpentry (R)
- Software Carpentry (Python)
- Data Carpentry Workshops
- Data Carpentry – 1st Annual Data Symposium
- Instructor Training
- Software Carpentry
- Multi-Week Data Carpentry

The Carpentries @ UF

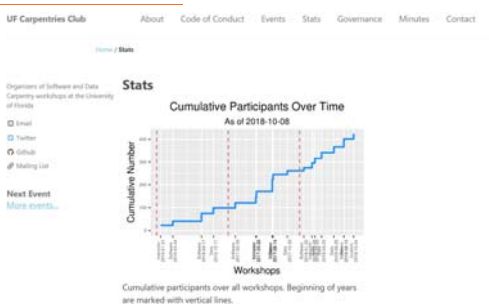


Fig. 2 – UF Carpentries Club Stats (https://www.uf-carpentries.org/stats/)

UF Data Science and Informatics (DSI)

UF DSI Focus for 2018

- ❑"Learning and teaching as fast as we can;"
- ❑"Creating a data science community" – DSI President
- ➢UF DSI is a recognized UF undergraduate student organization
  - ➢ http://www.dsiufl.org/

Developing new workshops

- ✓Use **Jupyter notebooks** and the anaconda distribution
- ✓Use **R Studio** for **R workshops**
- ✓Use **MySQL** for **SQL workshops**
- ✓Funded by UFII; collaborating with UF Libraries for a data science competition (spring 2019)
- ✓UFII fellows and students attend or help at workshops
- ✓Collecting DSI data for analysis for use by the Carpentries @ UF
- ❖Designing pathways a challenge

Academic Research Consulting & Services (ARCS)

- ❑Data Management
- ❑Publishing and Archiving
- ❑Research Metrics
- ❑Spatial Information Services
- ❑Statistical and Data Analysis
- ❑Visualization (2D, 3d, Virtual/Augmented Reality)
- ✓Bioinformatics Librarian
- ✓Data Management Librarian
- ✓Informatics Librarian
- ✓Geospatial Librarian
- ✓Physical Sciences Librarian
- ✓Political Science Librarian
- ✓University Librarian

UF ARCS

UF/IFAS USDA-NIFA funded collaborative research projects resulting in subawards to UF Libraries:

1. **SmartPath -- Grower-directed convergence of nanotechnology and smart decision analytics for irrigation water quality management related to pathogens** (subaward cash: $57,439)
2. **FACETS -- Floridan Acquifer Collaborative Engagement for Sustainability** (subaward cash: $10,376)
3. **Recovery and Resilience of Oyster Reefs in the Big Bend of Florida** (subaward cash to Libraries: $73,673)

Acknowledgements

- Matthew Collins, Technical Operations Manager, UF Advanced Computing and Information Systems Lab, ECE
- Tyler Richards, President of UF Data Science and Informatics

References

- ARCS. (2018). Academic Research Consulting & Services. http://arcs.uflib.ufl.edu/
- Davis, M. C., Challenger, R., Jayewardene, D. N. W., & Clegg, C. W. (2014). Advancing socio-technical systems thinking: A call for bravery. *Applied Ergonomics*, *45*(2 Part A), 171–180. https://doi.org/10.1016/j.apergo.2013.02.009
- UF Carpentries Club. (2018). https://www.uf-carpentries.org/; https://github.com/UF-Carpentry
- UF DSI. (2018). UF Data Science and Informatics http://www.dsiufl.org/
- UF Research Computing. (2018). https://www.rc.ufl.edu/

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

Measuring Reuse of Institutionally-Hosted Grey Literature

Elizabeth Kelly, Loyola University New Orleans;
Caroline Muglia, University of Southern California;
Genya O'Gara, Virtual Library of Virginia**;**
Ayla Stein Kenfield, University of Illinois at Urbana–Champaign**;**
Santi Thompson, University of Houston**;**
Liz Woolcott, Utah State University, United States

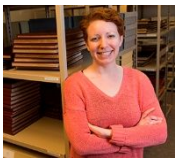
Grey literature is often hosted in digital libraries such as institutional repositories (IR) and managed by digital librarians who are tasked with proving the value of these collections. Newly developed multimedia COUNTER standards for IRs help guide collection and analysis of standardized use data but lack qualitative or storytelling measures that can add much needed nuance to assessment. Content reuse, or how often and in what ways digital library materials are utilized and repurposed, can bridge this gap for IRs.

The research team for Developing a Framework for Measuring Reuse of Digital Objects received a $70,850 grant from the Institute of Museum and Library Services (LG-73-17-0002-17) to address this problem. The grant's primary deliverables focused on the creation and application of an in-depth needs assessment of the digital library community to determine desired functionality of a reuse assessment toolkit. The research team's outputs include the development of well-defined requirements and use cases which serve as the building blocks for an assessment toolkit that goes beyond use and focuses on transformation.

The researchers employed surveys, focus groups, and data tagging and analysis to assess digital library needs for measuring reuse. Specific data points regarding reuse of IR materials were extracted and analyzed for this conference presentation. Measuring reuse of IR materials has emerged as a complex issue. Potential use cases thus far include: collecting data on what content was not reusable in order to develop deselection criteria; mapping dataset reuse to data management plans; assessing use and reuse of IR materials by grant funding agencies; and faulty or missing citations in electronic theses and dissertations. A broader concern emerged regarding weighting types of reuse differently depending on the type of digital collection as well as the mission and priority of the hosting institution. Participants identified lack of best practices, documented workflows, assessment training, and staffing as the greatest barriers to assessing reuse.

As the grant ends in June 2018, the research team has turned its focus to synthesizing the results of the needs assessment to address the challenges of measuring reuse. This presentation explores the findings of this research, specifically its impacts on Grey Literature hosted in IRs.

Bionotes



Elizabeth Joan Kelly, Digital Programs Coordinator at Loyola University New Orleans, manages digitization activities for Special Collections & Archives and is also responsible for collecting, maintaining, and assessing usage data for the library's digitized collections. Kelly publishes and presents on archives, digital library assessment, and library pedagogy, and co-founded the Digital Library Federation Digital Library Pedagogy group.
ORCID iD https://orcid.org/0000-0002-7306-3331

Caroline Muglia, Head of Resource Sharing and Collection Assessment Librarian at the University of Southern California (USC), manages the InterLibrary Loan and Document Delivery department and leads the collection assessment efforts for the Library system. In this capacity, she is responsible for qualitative and quantitative assessment and evaluation of all resources, the return on investment, and ways in which the library resources support research at the institution. Her current research interests include collection assessment, open education resources (OER), and streaming media opportunities in libraries.

Bionotes (CONTINUED)

Genya O'Gara is the Associate Director of the Virtual Library of Virginia (VIVA), a consortium of 72 academic libraries. In this position she implements consortial projects, coordinates assessment, develops collection management workflows, negotiates on behalf of members, supports committees and working groups, and assists in the preparation and management of consortial grants. She publishes and presents on emerging models of content development and assessment, with a focus on digital collections, scholarly publishing, and collaborative collection development.

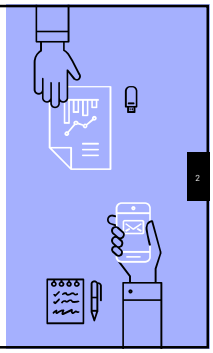
Ayla Stein Kenfield, Metadata Librarian at University of Illinois at Urbana-Champaign (UIUC). She supports the metadata needs for scholarly communication, data curation, and preservation in the Library. She has published and presented on digital repository evaluation, metadata development for data repositories, and digital library system migration. Her research interests include digital repositories; metadata and linked data; and the place of metadata in critical librarianship.
ORCID iD https://orcid.org/0000-0002-6829-221X

Santi Thompson, Head of Digital Research Services at the University of Houston (UH), serves as Primary Investigator for the grant. At UH Santi develops policies and workflows for the digital components of scholarly communications, including digital research support and digital repositories. He publishes on the assessment of digital repository metadata, software, and content reuse. Santi is currently the co-facilitator of the DLF AIG.
ORCID iD https://orcid.org/0000-0002-0337-6439

Liz Woolcott, Head of Cataloging and Metadata Services at Utah State University, manages the MARC and non-MARC metadata creation of the University Libraries and is the co-founder of the Library Workflow Exchange. She publishes and presents on workflow and assessment strategies for library technical services, innovative collaboration models, the impact of organizational structures on library work, creating strategic partnerships for libraries, and building consortial consensus for metadata standards.

Presentation Outline

I. Project Background
II. Data Collection Overview
III. Focus Group Sessions
IV. Use Cases - Grey Literature
V. Follow Up Survey
VI. More Information
VII. Next Steps



DLF

Assessment Interest Group (AIG)
Surveying the Landscape: Use and Usability Assessment of Digital Libraries (2015) white paper

INSTITUTE of Museum and Library SERVICES

2017 National Leadership Grant (LG-73-17-0002-17)
July 01, 2017 – June 30, 2018

Project Background

3

#digreuse
Project Team

reuse.diglib.org

Santi Thompson
University of Houston
sathomp3@central.uh.edu | @SantiThompson

Elizabeth J. Kelly
Loyola University, New Orleans
ejkelly@loyno.edu | @ElizabethJelly

Caroline Muglia
University of Southern California
muglia@usc.edu

Genya O'Gara
Virtual Library of Virginia, George Mason University
gogara@gmu.edu | @genyao

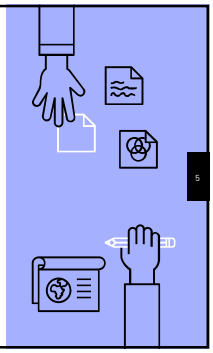
Ayla Stein Kenfield
University of Illinois Urbana-Champaign
astein@illinois.edu | @TheStacksCat

Liz Woolcott
Utah State University
liz.woolcott@usu.edu | @lizwoolcott

4

Assessing Use vs. Reuse

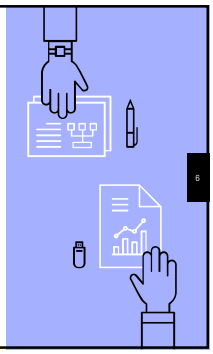
Use	Reuse
Discovering and browsing objects in a digital repository, often described as "downloads" or "clicks," without knowing the specific context for the use.	How often and in what ways digital repository materials are utilized and/or repurposed, knowing the context of the use.
Ex. Web analytics; CMS supplied analytics	*Ex. Social media use; reverse image lookup*



5

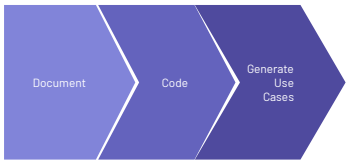
Data Collection Methods

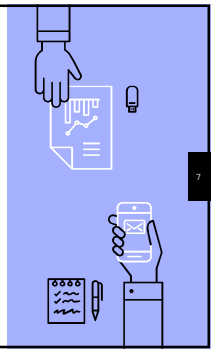
1. Initial Survey
2. Focus groups
   a. In-person
   b. Virtual
3. Follow-up survey



6

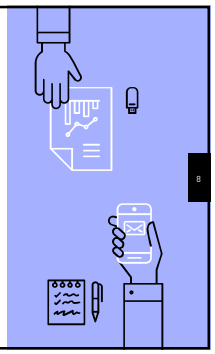
Focus Group Analysis





Examples of Reuse (so far...)

- Map dataset reuse to DMPs
- Citing digital objects in scholarly works, Wikipedia
- Transform data to stories/reports for various audiences
- Track non-scholarly reuse (genealogy, historic preservation, etc.)
- Reuse student data for curriculum develop.
- Blogs/web publishing
- Commentary/ annotation
- Auto-include rights statements
- Analog data transformed to digital datasets
- institutional history
- digital exhibits
- archiving websites
- documentary footage and/or research
- Database of historical vital records + connect to living people.
- presentations
- social media datasets
- measure grey lit. reuse w/ altmetrics
- Web analytics to track grey lit. URLs
- Track uses from aggregation services



Use Cases

Data Collection, Analysis, and Reporting

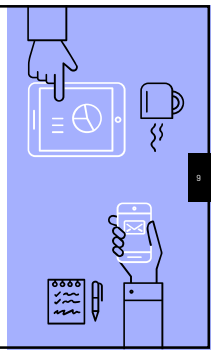
Use cases that address how reuse data is being assembled, interpreted, and communicated to various audiences and stakeholders.

Privacy, Rights Management, and Ethics

Use cases that address the assessment of reuse through legal, ethical, and culturally-aware perspectives

Collection Development

Use cases that address how to assess aspects of a digital collection based on reuse data.



1. Understanding how content is being reused in a variety of contexts by various audiences (social media, classrooms, scholarly works, genealogy, digital humanities, etc.)
1. Tell stories of impact with the reuse data that has been captured and tailor it to specific audiences/stakeholders
1. Assesses quality and quantity of items reused to inform digitization projects and priorities
1. Enable/encourage attribution of materials from collection in various reuse contexts, including through sharing and reposting on social media, integration into classroom instruction, citing in scholarly works, or through non-academic avenues
1. Know and understand digital repository users without violating user privacy

Follow-up Survey: Prioritized Use Cases

10

weighted and averaged (n=218, range 3.01-4.37)

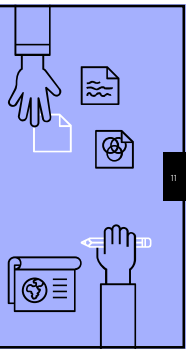
Next Steps

Spread the word

- *Performance Measurements & Metrics*, " Barriers and Solutions to Assessing Digital Library Reuse: Preminimary Findings"
- ARL Assessment Conference

Apply for new grant

Second grant to **build** the toolkit based on use cases & functional requirements



11

THANKS!

Any questions?

reuse.diglib.org #digreuse

12

Credits:
Presentation template by SlidesCarnival
Photographs by Unsplash

Librarians' Role in GAO Reports

Meg Tulloch, Colleen Candrl, and Leia Dickerson
U.S. Government Accountability Office, GAO, United States

Problem: In a time when information is freely available, what role is there for librarians in the research process, most specifically the U.S. Government Accountability's research process that leads to GAO Reports?

Method: Librarians play several roles in the research process that involve finding, collecting, and analyzing large sets of information and/or data. The librarians at the GAO play a formal role in the research that supports the findings in GAO reports. As determined by the engagement teams in conjunction with methodologists and librarians, literature reviews with specific criteria can be used to provide evidence for findings. Literature and background searches also support the research behind the reports. This presentation/paper will provide examples of the three types of searching—literature review, literature search, and background searching—and their role in the GAO process. It will also discuss our method for assessing Web-based sources of information.

Result: Through literature reviews, librarians can produce evidence for GAO Reports. Literature and background searches also support the research behind the reports.

Bionote

Meg Tulloch is Executive Director of the Federal Library and Information Network (FEDLINK), Library of Congress. She is the former Library Director of the National Defense University Libraries in Washington, D.C. and Norfolk, Virginia. Previously, she was the Europe Region Librarian for the U.S. Army and oversaw 26 libraries in four different countries. She has also worked as a librarian at Vanderbilt University's Walker Management Library and Kutztown University of Pennsylvania's Rohrbach Library. Much of her career has focused on how technology can assist the researcher through digital library tools, using digital materials. Additionally, Meg taught "Introduction to Poetry Writing" at the University of Virginia while a graduate student there. She holds a Masters in Library and Information Science, a Masters in Fine Arts in poetry writing, and a Bachelors in American Literature. She is currently pursuing a Doctorate of Liberal Studies from Georgetown University. Her dissertation will explore fragmented twenty-first century literature. Email: mtulloch@loc.gov

GAO

Librarians' Role in GAO Reports

Presenter: Meg Tulloch
Contributors: Colleen Candrl, Leia Dickerson
December 3, 2018

The opinions expressed in this presentation are the author's own and do not necessarily reflect the views of the U.S. Government Accountability Office.

Meg Tulloch, 202-512-8535 | TullochH@gao.gov | https://www.gao.gov/ Page 1

GAO

About GAO

- GAO – The Congressional Watchdog – YouTube

- https://www.gao.gov/about/what-gao-is/performance/

Page 2

GAO

GAO Stats and Background

Branch: Legislative

People: 3,000 employees

Locations: 70% in HQ in DC with field offices and staff in 10 other major cities

Organization: 14 Mission teams plus internal support offices (Congressional Relations, Public Affairs)

Work: Audits of Government Programs

Authority:

- GAO's work is done at the request of congressional committees or subcommittees;
- Mandated by public laws or committee reports;
- GAO also undertakes research under the authority of the Comptroller General.

Page 3

GAO

Library Services Situated Within GAO:

Organization	
Executive Committee	
Comptroller General (CG)	Chief Operating Officer (COO)
Chief Administrative Officer / CFO	General Counsel (GC)
Staff Offices	
Audit Policy and Quality Assurance (APQA)	Continuous Process Improvement (CPI)
Congressional Relations (CR)	Opportunity & Inclusiveness (OI)
General Counsel (GC)	Office of Public Affairs (OPA)
Inspector General (IG)	Strategic Planning & External Liaison (SPEL)
Chief Administrative Office	
CAO Immediate Office (CAO)	Infrastructure Operations (IO)
Field Operations Home	Information Systems and Technology Services (ISTS)
Financial Management and Business Operations (FMBO)	Learning Center (LC)
Human Capital (HCO)	Professional Development Program (PDP)
Mission Teams	
Applied Research and Methods (ARM)	Forensic Audits and Investigative Service (FAIS)
Contracting and National Security Acquisitions (CNSA)	Health Care Team (HC)
Defense Capabilities and Management (DCM)	Homeland Security and Justice (HSJ)
Education, Workforce, and Income Security (EWIS)	Information Technology (IT)
Financial Management and Assurance (FMA)	International Affairs and Trade (IAT)
Financial Markets and Community Investment (FMCI)	Natural Resources & Environment (NRE)
	Physical Infrastructure (PI)
	Strategic Issues (SI)

Page 4

GAO

Direct Support for GAO Reports

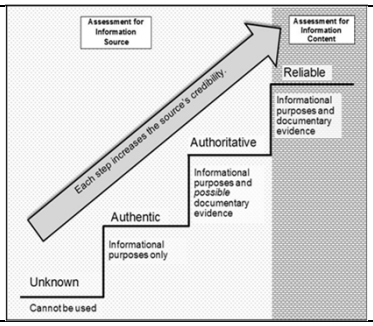
Direct Support to the Mission Teams Responsible for GAO Reports

- **Stakeholders -** Developing teams scope and methodology (Pre-Design Matrix)
- **Literature Searches**
- **Background Searches**
- **Acquisitions** (Large and Small $$$$)
- **Interlibrary Loans and Document Delivery**
- **Data Reliability:** See Guidelines for Assessing Web-based Sources of Information.

Page 5

GAO

Guidelines for Assessing Web-based Sources of Information



Page 6

GAO

Literature Search and Review

Two Purposes for Reviewing Literature:
- Background
- Support engagement objectives

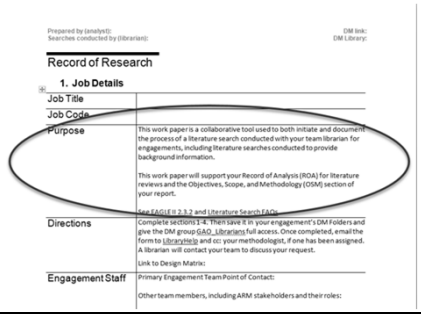
Process:
- Record of Research
- Citation Manager (Refworks.com)
- Data Collection Instrument (DCI)
- Objectives, Scope, and Methodology Statement

Guidance:
- EAGLE 2.3.2
- Using Studies Conducted by Outside Researchers for GAO Engagements
- Guidelines for Assessing Web-based Sources of Information
- Objectives, Scope, and Methodology (OSM) Guidance

Page 7

GAO

Record of Research



Prepared by (analyst):
Searches conducted by (librarian):

DM link:
DM Library:

Record of Research

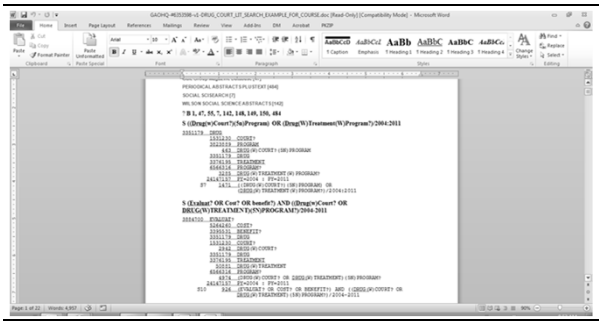
1. Job Details

Job Title	
Job Code	
Purpose	This work paper is a collaborative tool used to both initiate and document the process of a literature search conducted with your team librarian for engagements, including literature searches conducted to provide background information. This work paper will support your Record of Analysis (ROA) for literature reviews and the Objectives, Scope, and Methodology (OSM) section of your report. See EAGLE 2.3.2 and Literature Search FAQs
Directions	Complete sections 1-4. Then save it in your engagement's DM Folders and give the DM group GAO_Librarians full access. Once completed, email the form to LibraryHelp and cc: your methodologist, if one has been assigned. A librarian will contact your team to discuss your request. Link to Design Matrix:
Engagement Staff	Primary Engagement Team Point of Contact: Other team members, including ARM stakeholders and their roles:

Page 8

GAO

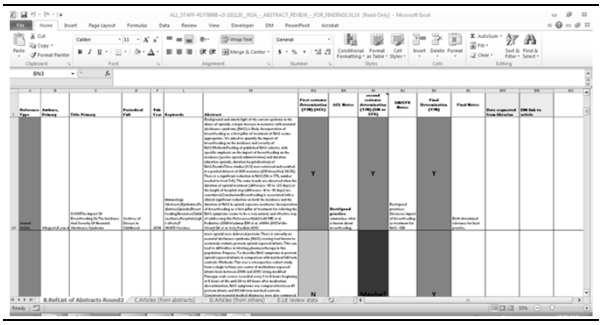
Record of Research



Page 9

GAO

Data Collection Instrument (DCI)



Page 10

GAO

PRELIMINARY

GAO-10-344: CHILD CARE

https://www.gao.gov/products/GAO-10-344

United States Government Accountability Office

GAO Report to Congressional Requesters

May 2010

CHILD CARE

Multiple Factors Could Have Contributed to the Recent Decline in the Number of Children Whose Families Receive Subsidies

Page 11

GAO

Objectives, Scope, and Methodology Statement (OSM)

Review of Literature on Child Care Subsidies

To determine what is known about the impact of child care subsidies on employment, we conducted a literature search for studies that analyzed relationships between child care subsidies or changes in child care costs and employment outcomes. To identify existing studies from peer-reviewed journals, we conducted searches of various databases, such as EconLit, ProQuest, PolicyFile, and Social SciSearch. We also asked all of the external researchers that we interviewed to recommend additional studies. From these sources, we identified 31 studies that appeared in peer-reviewed journals between 1995 and August 2009 and were relevant to our research objective on the effect of child care subsidies on employment outcomes. We performed these searches and identified articles from June 2009 to October 2009.

To assess the methodological quality of the selected studies, we obtained information about each study being evaluated and about the features of the evaluation methodology. We based our data collection and assessments on generally accepted social science standards. We conducted an extensive literature review, examined summary level information about each piece of literature, and then from this review, identified articles that were germane to our report. We then evaluated the methods used in the research, eliminated some research if we felt the methods were not appropriate or rigorous, and then summarized the research findings. In addition, for articles directly cited in the report, we performed an initial in-depth review of the findings and methods, and then a GAO economist performed a secondary review and confirmed our reported analysis of the finding. As a result, the 31 studies that we selected for our review met our criteria for methodological quality. We supplemented our synthesis by interviewing four of these studies' authors. We also conducted an interview with an official at the Office of Planning, Research and Evaluation within the Administration for Children and Families.

Page 12

Library Contributions to Research Data Management and Open Access at the National Institute of Standards and Technology

Regina L. Avila, National Institute of Standards and Technology, USA

Over the last several years National Institute of Standards and Technology (NIST) Librarians were important players in creating and promoting services to enhance the management of research data. The NIST Research Library has played a central role in educating researchers, assisting in the creation of tools to manage data, and promoting practices to enhance data curation – beginning with their membership on the earliest agency committee discussing data management to most recently being awarded NIST's highest honor for making data publicly available. This poster will illustrate the many ways the NIST Research Library promoted data management practices and public access to data, including:

- Co-authorship of The NIST Plan for Providing Public Access to Results of Federally Funded Research
- Co-creation of NIST's first tool to create Data Management Plans
- Assistance in creating official data management policies and practices
- Collaborations with the NIST Material Measurement Lab and the Engineering Lab to promote data curation and data publishing
- Assistance in creation of applications for reviewing and sharing data
- Contributions to the creation of a data repository with preservation capabilities

Since 2010, when NIST formed its first committee to review research data management, the NIST Research Library has consistently contributed to this work and continues to collaborate with researchers and administration to improve data management practices and increase access to NIST data.

Bionote

Regina Avila is a Digital Services Librarian and Data Librarian at the National Institute of Standards and Technology. She joined NIST in 2009 to implement the agency's NIST Digital Archives, leading a team in the digitization and preservation of the agency's publications. Since then her duties have included leading a library team to investigate research data management services. As a Data Librarian, she successfully collaborated with NIST's IT department and laboratories in curating and publishing research data – including the development of tools to create data management plans, the NIST enterprise data inventory, and data documentation. She is co-chair of the Data Curation Discussion Group within CENDI, an interagency group of senior Scientific and Technical Information (STI) managers from 14 federal agencies. Email: regina.avila@nist.gov

Metadata Drives Special Collections Research: A Case Study of the University of Florida Thesis and Dissertations Collection

Chelsea Dinsmore, University of Florida; George A. Smathers Libraries, USA
Daniel Vasicek, Access Innovations, Inc. USA

In modern library systems, access to collections is heavily dependent on effective metadata. Metadata is made up of many parts and even "Subject metadata" can be divided into several types. To test the efficacy of different kinds of metadata used by the research community at the University of Florida we applied two additional Subject metadata access points to the thesis and dissertations collections. Existing Subject metadata came from the Library of Congress subject headings. After analyzing the actual subject related headings versus the administrative information, UF added thesaurus terms from the JSTOR Thesaurus and from the Geographic Place Name thesaurus to enhance the potential subject access based on the OCR'd full text of each document. The accuracy in test searches using the added subject headings compared to the original LCSH headings was measured. Search queries against LCSH headings only; LCSH and title; Title and JSTOR headings; Title, JSTOR, Geographic headings; LCSH, JSTOR and Geographic; and finally all subject access plus title points were measured to compare efficiency of retrieval on the full corpus against a standard query set. This poster presents the results of that study as a benchmark for future search implementations and provides the basis for the planned use of subject and geographic indexing on all UF Collections.

Bionotes

Chelsea Dinsmore serves as the Director of Digital Production Services for the University of Florida George A. Smathers Libraries and is the Technical Director for the Digital Library of the Caribbean (dLOC). Since 2009, she has organized and managed digital projects within the government documents community, including the creation of regional federal depository Centers of Excellence (COE) for the Panama Canal and the National Recovery Administration collections. She is currently involved in research to improve the accessibility of digital collections by rehabilitating legacy metadata. Ms. Dinsmore holds an MLIS from the University of Texas at Austin and an MS in History from the University of Florida.

Daniel Vasicek is a Data Scientist working with Access Innovations to apply computer science algorithms to the indexing of documents.

"M@dNotes Project" Giving colour to Grey Literature in General Practice: A collaborative GPs knowledge network

Miguel Pizzanelli MD, MSc, Melissa Resnick, Elena Cardillo, and Marc Jamoulle
M@dNotes / NotasL@cas, GPs Collaborative international knowledge network, Uruguay

Sharing the results of research and scientific production is crucial for the survival of all disciplines. Health information is becoming uncertain and powerful economic interests disrupting medical information has triggered a loss of credibility. M@dNotes project (in Spanish NotasL@cas), is the by-product of non-profit collaborative international knowledge network of general practitioners, family physicians, and healthcare workers. One of the aims of this network is to discuss the validity of the information available on sensitive subjects and build a knowledge exchange network inside a community of practice. The project was born because we realized that more than 95% of the postgraduate and pregraduate research/reports/ scholar works (monographic works) to accomplish or finished one part of the training were finally lost and unpublished. Most of them come from interesting "minds" and are interesting topics to improve the applying of evidence in medical healthcare. From this perspective, M@dNotes is a unique local platform to rescue knowledge and support exchange and collaboration. Perhaps is isolated due to the circumstance of being restricted to few countries or a small group of practice. Nevertheless, we are trying to improve our qualifications in order to link M@dNotes to an open access network. It has been created to share different materials excluded from traditional academic publishing and commercial distribution channels: monographic, commented articles, files with references searches. These materials considered "grey literature" are the product of intense learning process and has been produced to achieve a qualification, be presented in a conference or were collected by users spontaneously through the web searches, with a knowledge purpose. Many of them qualified with an appropriate level because were submitted to a peer review or judged by a scientific committee to get an approval. The project includes an indexing method (Q Codes and ICPC) to manage and retrieve the materials received. Q-Codes are intended for indexing Family medicine related documentation jointly with the International Classification of Primary Care (ICPC). The project provides a particular folder to upload and share the documents. A web blog is the dissemination tool or friendly interface to share and retrieve the documents shared in the network. Through a facility in the cloud, it is possible to access an excel file with all the materials shared, codes assigned and the hyperlink to get a pdf file of each material.

Bionotes

Miguel Pizzanelli, MD, MSc. - My name is Miguel Pizzanelli Báez. I was born in Montevideo in 1962. With Virginia my wife and dear partner, we share raising three children. At this moment we live in Florida, Uruguay. Since 1996 I spent almost all my medical practice time in small rural areas. For 7 years (from 2003 to 2010) we had the experience of living and working in a small rural village of 1500 inhabitants. I have varied interests, reading, I try to play several musical instruments in a self-taught way. Hobbies: music, trekking, photography, web-blogs editor. I am general practitioner (family and community medicine) from 2003. I was part of the first generation of family and community medicine residents trained in Uruguay. I call this the zero generation (remembering of hard times that passed). I use to disseminate contents in various topics: quaternary prevention, rural medicine, critical thinking development. Quaternary prevention is a concept that defines an attitude ethically center oriented to provide health care focusing on persons trying to share health decisions with them in order to avoid overmedicalization. Since 2012 I began to actively participate in the society of family and community medicine in Uruguay and from that place in CIMF / WONCA. My role leading dissemination and applied of quaternary prevention concept pushed me to lead quaternary prevention working groups, first in my country later in Iberoamerican region and now in WONCA. My interest in classification and systematic terminologies makes me accept the invitation to participate in WONCA International Classification Committee in the quality of associate member from November 2014 up to date. Since 2008 we develop research focus on Barbara Starfield's Primary Care Assessment Tool in Uruguay. I participate actively in national regional and international CIMF WONCA Conferences (Praga 2013, Montevideo 2015, and Rio de Janeiro 2016). I think we need to fight both an individual and collective fight. The Individual fight to set collective interests over personal ones. Only through the collective work of all the family doctors and communities together all over the world, we will achieve "real" Primary Care: comprehensive health care, equity, and people-centered health care, focus on health better than illness, making reality the utopia of health for all in a better world.

Published electronic media are becoming Grey

Yui Kumazaki, Satoru Suzuki, Masashi Kanazawa, Katsuhiko Kunii,
Minoru Yonezawa, and Keizo Itabashi
Japan Atomic Energy Agency, JAEA, Japan

The library of the Japan Atomic Energy Agency (hereinafter, this is called "JAEA Library") focuses on collecting conference proceedings and technical reports which are very important and referred materials for our users in the field of nuclear science and technology. There are an increasing number of cases that they are published in electronic media such as CD-ROM and flash memories. There are problems as to manage such electronic media in a library. For example, the shape of them are quite different from ordinary books and the size is very small and they would be easily lost if we manage them in improper way. So we use the plastic case of the same shape for them and store on the book shelves in order to manage easily and to avoid lost. Largest problems for electronic media would be the long-term preservation and permanent access.

Appropriate equipment with software in PCs are necessary to read these publications in electronic media. However, it is quite common such electronic media are used to be updated or changeable on specifications and that the older become unavailable in a short period. Lifetime of electronic media are much shorter compared with paper and/or microfilm.

For example, JAEA Library hold books and technical reports of 1990s in electronic media. Some of them have already been unavailable because they have no appropriate software to read. In order to read such electronic media that cannot be played on the present PCs, JAEA Library preserve PCs of Windows 95, 98, etc which have already been out of supports.

In this regard, the JAEA library is struggling to preserve electronic media including related PCs and software for a long-term. One of the effective countermeasure should be preserving PCs and software to browse, emulation and migration to another media. However, our countermeasure is not good enough for users and preservation.

The authors would like to present JAEA Library's current activities on long-term preservation and use of electronic media, and it comes to the conclusion the fact that even published electronic media are becoming grey literature in certain environment.

Bionote

Kumazaki Yui is a librarian of the Japan Atomic Energy Agency. Email: kumazaki.yui@jaea.go.jp

Published electronic media are becoming Grey.

Yui Kumazaki, Satoru Suzuki, Masashi Kanazawa, Katsuhiko Kunii, Minoru Yonezawa, Keizo Itabashi
Japan Atomic Energy Agency

About the JAEA Library

- One of the largest nuclear information centers in Japan
- INIS (International Nuclear Information System) national center of Japan
- Fukushima Nuclear Accident Archive (FNAA)
- Publishing technical reports (JAEA Reports)
- Dissemination of JAEA research results

FNAA JOPSS JAEA Originated Papers Searching System

2

Background

- Focuses on collecting conference proceedings
  - Very important and referred materials in the fields of nuclear science and technology
  - **76%** of JAEA achievements announced at the conferences (and papers) (2015/04/01-2018/03/31)
  - Sometimes difficult to be shared
  - To improve the accessibility to the proceedings, which **were Grey**, by collecting

3

Background

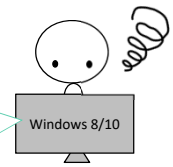
Paper → **Electronic media : CD/DVD-ROM, flash memories**

4

Problem of electronic media

- Having made the proceedings more and further **Grey and useless**
- Due to:
  - Lifetime of electronic media and playback devices are short
  - Changes in standards/specifications of media and playback devices

"Not supported"



5

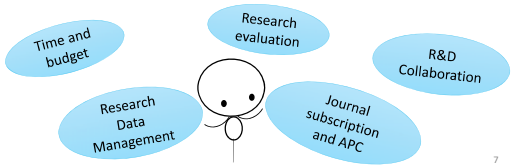
Problem of electronic media

Windows 95/98
Purchased in 2000
Difficult to print.

6

Possible countermeasures, But …

- Long-time preservation and Ensuring accessibility
  - Emulation
  - Migration as conversion of file format etc.
  - Maintenance of the regeneration environment
- But we do not have the enough resources



7

What is important for users?

- After all, making up our mind,
  - The users should be able to read articles!
  - The users should be able to print them!
- We prioritized what we can **read and print**
  - By using Windows10(32bit)

8

Thank you for your kind attention

- How many electronic media are there around you?
- How are you dealing with?

- Would you tell me what you are in trouble with?

9

GreyGuide

Repository and Portal to Good Practices and Resources in Grey Literature

2013 - 2018
5 Years of collaborative effort *sharing Knowledge and Data on Grey Literature*

Welcome to the GreyGuide, your point of access to Grey Literature and Open Access Resources

The GreyGuide - Repository and Portal to Good Practices and Resources in Grey Literature was launched in 2013 as a collaborative effort between GreyNet International and ISTI-CNR, Pisa, Italy.

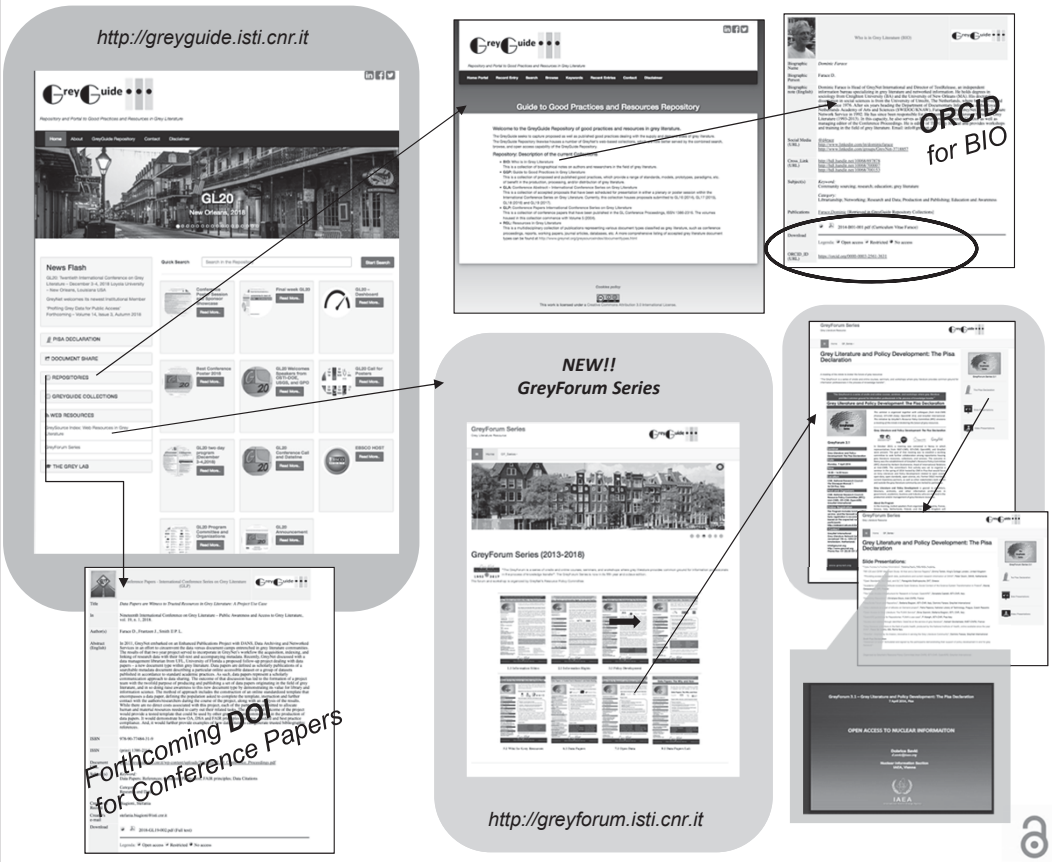
2013-2017
- Repository
- Portal
- Document share

2018
-Grey Forum Series Portal
- DOI for Conference Papers
- ORCID for BIO Notes

The GreyGuide manages Open Source Repositories and provides a unique resource in the field of grey literature that is long awaited and which responds to the information needs of a diverse, international grey literature community adhering to Open Science guiding principles.

This year activities dealing with the GreyGuide have focused on
- upgrading the new version of its Portal;
- designing **a new portal for the GreyForum Series** enabling access to material produced by the speakers;
- including an **ORCID metadata field** in the BIO Collection in the GreyGuide Repository;
- including a **DOI metadata field** for GL-Conference Papers in the GreyGuide Repository *(Work in Progress)*

In 2019, GreyGuide will embark on a 7-Point Plan of Action seen as its roadmap for the coming years in serving GreyNet International as well as other communities of practice in the field of grey literature.



PARTNERS

GreyNet
http://www.greynet.org/

CNR ISTI
http://www.isti.cnr.it

CONTENT PROVIDER

Dominic Farace
GreyNet International
Netherlands

SYSTEM MANAGER

Stefania Biagioni
ISTI-CNR
NeMIS Lab
Pisa, Italy

TECHNICAL DEVELOPER

Carlo Carlesi
ISTI-CNR
NeMIS Lab
Pisa, Italy

COLLECTIONS

- BIO: Who is in Grey Literature
- GGP: Guide to Good Practices
- GLA: Conference Proposals
- GLP: Conference Papers
- RGL: Resource in Grey Literature

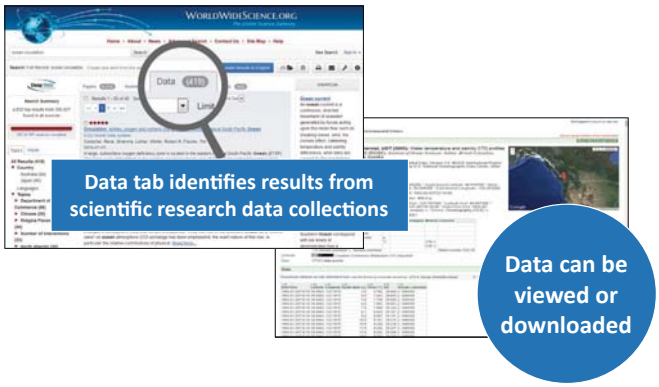
Remember to endorse The Pisa Declaration

WorldWideScience.org

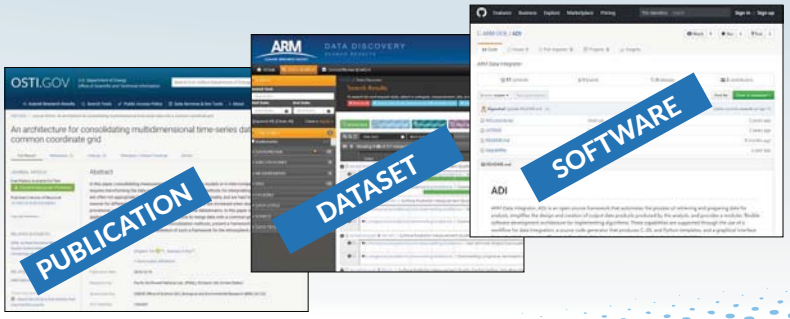
An International Partnership Supporting Open Science

- Simultaneously explore over 100 national and international scientific databases and portals from 75 countries
- Search information in textual, multimedia, software, and scientific data formats
- Eliminate language barriers through multilingual translations across ten languages

Search and Access Scientific Research Data



Modern Science Demands Reproducibility and Open Access to Publications, Datasets, and Software



WorldWide ScienceAlliance

Operating Agent:

OSTI.GOV

WorldWideScience.org Operating Agent
Point of Contact: Lorrie Johnson, JohnsonL@osti.gov

WorldWideScience.org: An International Partnership Supporting Open Science

Lorrie A. Johnson,
U.S. Department of Energy Office of Scientific and Technical Information, USA

The WorldWideScience Alliance is a strategic partnership, comprised of national and international libraries and information centers, whose goal is to eliminate barriers to finding and sharing scientific and technical information across national boundaries. The Alliance provides the governance structure and sets the direction for WorldWideScience.org, a federated search portal. WorldWideScience.org offers users the ability to simultaneously search, in real time, over 100 scientific and technical databases from more than 70 countries. Search results from the various databases, which include text, multimedia, software, and research data, are relevance-ranked and a consolidated results list is presented to the user. Multilingual translations capabilities are available for ten languages, which makes scholarly material more accessible worldwide. The ability to search data and software collections within WorldWideScience.org addresses many of the challenges associated with discoverability of research datasets and scientific software. WorldWideScience.org enables users to receive data and software results in a separate results tab, and upon selecting a specific result, users will typically be directed to a landing page, which in turn makes the data or software accessible for viewing or downloading. The inclusion of data and software collections in WorldWideScience.org supports interlinking between publications, datasets, and software. As part of the broader public access movement among government research funders in many countries, WorldWideScience.org further expands access to R&D results across the full research lifecycle, and ultimately contributes to increased scientific collaboration and progress.

National Repository of Grey Literature (NRGL)

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

Národní technická knihovna
National Library of Technology

národní úložiště šedé literatury

Milestones of Grey Literature in the Czech Republic

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

This poster aims to take you through the all-important milestones and moments of Czech grey literature history. The history is still brief, but not poor at all. Until 2005, the National Library of Technology in Prague (NTK) was in charge of distributing grey literature data from the Czech Republic into the System for Information on Grey Literature in Europe (SIGLE). That year, SIGLE was discontinued, and that led to a project at NTK to build the Czech National Repository of Grey Literature (NUŠL) in order not to lose valuable documents created outside the main publishing stream. The preparatory phase took about two years, and in 2008, the project began led by NTK with financial support from the Ministry of Culture of the Czech Republic. This was one of the biggest milestones for Czech grey literature, with NTK's role shifting from contributor to central collector and leader of this topic in the Czech Republic. This year is the tenth anniversary since the kick off day of the project, and many things have changed since then, including software, features, design, people, and much more. The Seminar on Providing Access to Grey Literature that has been linked to the project from beginning has evolved as well. In 2014, it was transformed into an international Conference on Grey Literature and Repositories. This poster evaluates the progress of this repository and its efforts in collecting, preserving, sharing, providing accessibility to, and promoting grey literature in the Czech Republic. Past experiences lead us to consider the future of grey literature as in general as in case of NUŠL and maybe also its reassessment.

Bionotes

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

GreyHub: A Discovery Service for Grey Literature

Abe Lederman, Deep Web Technologies, USA

Deep Web Technologies (DWT), the company that built Explorit Everywhere!, the search engine that powers WorldWideScience.org and Science.gov, is creating GreyHub.org. The new site will serve as the one-stop access portal to the world's largest collection of grey literature. GreyHub will be a collaboration between DWT, GreyNet, and others in the Grey Lit community. The joint effort will create a catalog of potentially thousands of curated Grey Lit repositories. Explorit Everywhere! will provide real-time search access to some of the best repositories in the catalog. We expect to be able to search millions of Grey Lit reports, government publications, monographs and other resources. The collaboration will also work with a DOI Registration Agent and with Orcid.org to assign DOI/ORCIDs to resources as needed. These enhanced resources will become part of the GreyHub portal.

Bionote

Abe Lederman has 35 years of experience in computer software engineering, the last 25 years focused exclusively on developing leading-edge information retrieval products and solutions. He began his career as a Software Developer at Hewlett Packard and was then recruited in 1987 to become a founding member of Verity, a startup pioneer in the field of text retrieval where he developed the user interface for their Topic product. While consulting with Los Alamos National Laboratory (1994 - 2000) and later with the U.S. Department of Energy's Office of Scientific and Technical Information (OSTI), Abe came to realize the enormous potential for federated search technology to accelerate the diffusion of knowledge. He founded Deep Web Technologies in 2002. As CEO and CTO, Abe has pushed DWT to set the gold standard for Federated Search, pushing the limits of what is possible. Abe holds Bachelor of Science and Master of Science degrees in Computer Science from the Massachusetts Institute of Technology.

ORCID ID: https://orcid.org/0000-0002-9970-5590

Never Stop Asking Why. Data Tells the Story….and Libraries Help

Julia Gelfand, University of California, Irvine, USA
Anthony Lin, Irvine Valley College, USA

Data management requires special skills as does the acquisition, processing and incorporation of other forms of grey literature. The data science lifecycle parallels the scholarly publishing continuum and includes the infrastructure needs of storage, analysis, visualization and interpretation. Increasingly, the privacy and ethical concerns associated with data security are paramount in the public sphere. The information enterprise depends on understanding data to direct future initiatives. Libraries have demonstrated the capacity to absorb many publishing and administrative functions of academic scholarship including:

- attending to data privacy as it relates to user activity and patron records
- collecting and sourcing data from primary and secondary sources
- managing research data and organizing data metrics/analytics to determine various impacts
- leading data curation
- promoting data literacy at all levels of the academic ladder
- navigating big data inventories
- promoting open data
- establishing appropriate data repositories.

Data is often complicated, elusive, and not always intuitive. Developing programs and services that meet the needs of a data-intensive universe where evidence supports outcomes is an underlying basis for how data can tell a more accurate story that can lead to better informed decisions, predictions and hold greater value. The increased interest in how different data streams, such as twitter logs are influencing directions and new ideas reinforce how grey literature must support data in all its wrappings.
This presentation will focus on how data science and data management contribute to increasing value and how libraries and their more data centered missions are leading to transform the profession of a research librarian, and academic scholarship across the disciplines. If applying Jeannette Wing's "data for good" landscape, society benefits from more open access to data, better tools to understand it and ways to incorporate data in our efforts to reduce misinformation and to understand how data advances science and medicine, public policy, the arts and environment, and opens up new relationships with industry and commerce. With more machine learning, artificial intelligence contributing to how we simulate and share data, the possibilities for the stewardship and management of data only become less grey and more significant.

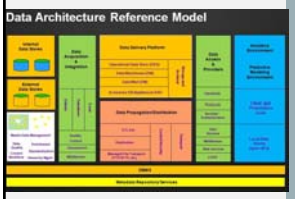
Bionotes

Julia M. 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

DATA TELLS THE STORY & LIBRARIES HELP

Julia Gelfand and Anthony Lin
University of California, Irvine and Irvine Valley College Libraries
20th International Grey Literature Conference 2018
New Orleans, LA, USA



Source: LRDC. "DBARCH Data Management Reference Model." Wikimedia Commons, Creative Commons Attribution, 28 Dec. 2009, commons.wikimedia.org/wiki/File:DBARCH_Data_Management_Reference_Model.JPG.

EFFECTIVE DATA MANAGEMENT PROMOTES

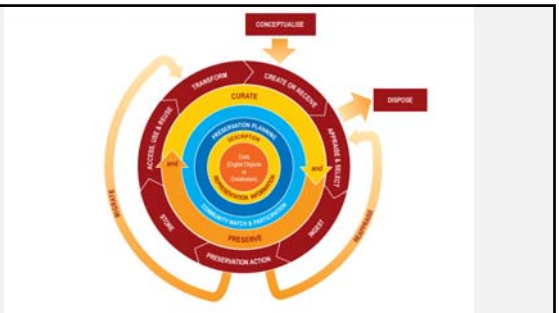
- Best Practices for Curation & Preservation
- Project Management
- Privacy
- Security
- Discovery & Access
- New Publishing Models
- Better Understanding & Appreciation for Grey Literature

TRENDS IN OPEN EVERYTHING

• Open Access • Open Science • Open Data • Open Publishing • Open Grey	Ensuring quality control via peer review leads to successful design, reproducibility, sharing and collaboration on multiple scales
• Data Literacy	Accurately Sourcing Data Implementing Data Management Plans Role of Repositories
• Developing New Skill Sets	Preservation & Stewardship Accompanying Data Sets for Reuse Cloud Solutions
• Data Management	Metadata Establishing Data Lifecycle

DATA ADDS VALUE

- Content is richer
- Evidence-based contributes to credibility & reproducibility
- Confirms scholarship with unique attributes
- Data reuse builds on established findings & norms
- Collected data lives on through repurposing – extends lifespan
- New ideas lead to innovation in methodologies and outcomes



Key elements of the DCC Curation Lifecycle Model

Source: The University of Edinburgh. "Key Elements of the DCC Curation Lifecycle Model." DCC , Creative Commons, 2008, www.dcc.ac.uk/resources/curation-lifecycle-model.

HOW LIBRARIES & LIBRARIANS HELP

- Process data
- Promote openness
- Serve as intermediary between researcher & findings
- Introduce appropriate tools & services to manipulate & manage data – data carpentry
- Advise researchers how to retain data management rights
- Host data in repositories for safekeeping & ongoing review
- Assist users in finding appropriate data for information queries

DATA ON THE GREY SPECTRUM: "DATA FOR GOOD"

As information elements have more searchable components the black to white continuum blends changing the hues to more AND less greyness. Data is central to understanding, describing and validating research outcomes. Data description is critical to information sharing in the ever-changing landscape.

Poster tells more of the story!

Documentary Film, TV Series, and Investigative Journalism as Grey Literature

Robert E. Noel, Indiana University Libraries, USA

Documentary film and news series (e.g. *Frontline* and the award winning long-standing news magazine *60 Minutes*) offer researchers and students a unique perspective; they are an encapsulation of current affairs. They offer a more consolidated perspective than traditional scholarly monographs do, by distilling, depicting, and describing difficult subject matter in a relatively short time period. Streaming video of these types of resources are available to academic research libraries from publishers such as Alexander St. Press, a ProQuest company.

What these episodes or pieces may lack in terms of comprehensiveness, they make up for in their ability to frame issues. John Grierson, an early 20th century documentary filmmaker, offered a definition of documentary as "creative treatment of actuality". Around that same time, American filmmaker Pare Lorentz defined documentary as "a factual film which is dramatic".

Extended news pieces and documentaries can be used to start conversations that may be more fully explored and researched in the undergraduate or graduate classroom. Further, the multidisciplinary nature of key issues seems to surface more readily with documentary film than traditional print media. While documentaries may strive to be objective, they are often controversial; they deliberately foment, and at times reflect the filmmakers' advocacy and bias. This poster offers an overview of important documentaries and series, and suggests how they may be suitable resources for courses on college campuses.

Examining Grey Literature Use, Citation, and Creation Practices of Faculty at a Large Institution

Kristen Cooper, Wanda Marsolek, Amy Riegelman, Shannon Farrell, and Julie Kelly
University of Minnesota Libraries, USA

Various disciplines, such as economics and forestry, depend heavily on grey literature, as is evidenced by the unique and extensive collections of grey literature in these subjects in university libraries. However, other fields' use of grey literature may vary extensively. This study was undertaken to determine the extent of use and creation of grey literature by faculty at the University of Minnesota. The research team consisted of five subject librarians. For the purpose of this research, grey literature was defined as works such as conference papers, working papers, technical reports, versions of articles submitted for publication, theses and dissertations, and government documents. The study aimed to discover 1) whether the library is meeting the needs of faculty who are using or creating grey literature, 2) what the differences are between disciplines and what type of literature faculty are using and creating, and 3) what resources or services might assist faculty in working with grey literature. The study design was mixed method and included an anonymous online survey and 60 minute recorded in-person interviews.. The survey found that faculty who use grey literature use many types, including conference papers, technical papers, and working papers. Reasons for use included locating descriptions of technical methods and staying current on trends. For faculty who indicated that they do not create or cite grey literature, reasons included not receiving credit for producing it and concerns about its quality. Preliminary results from the interviews and plans for qualitative analysis of transcripts will be discussed. The results of the study will potentially be used to help shape policy surrounding grey literature such as recruitment for the institutional repository.

Semantic Query Analysis from the Global Science Gateway

Sara Goggi, Gabriella Pardelli, Roberto Bartolini, and Monica Monachini, ILC-CNR, Italy
Stefania Biagioni and Carlo Carlesi, ISTI-CNR, Italy

Nowadays web portals play an essential role in searching and retrieving information in the several fields of knowledge: they are ever more technologically advanced and designed for supporting the storage of a huge amount of information in natural language originating from the queries launched by users worldwide.

A good example is given by the *WorldWideScience search engine:*

The database is available at <http://worldwidescience.org/>. It is based on a similar gateway, Science.gov, which is the major path to U.S. government science information, as it pulls together Web-based resources from various agencies. The information in the database is intended to be of high quality and authority, as well as the most current available from the participating countries in the Alliance, so users will find that the results will be more refined than those from a general search of Google. It covers the fields of medicine, agriculture, the environment, and energy, as well as basic sciences. Most of the information may be obtained free of charge (the database itself may be used free of charge) and is considered ''open domain.'' As of this writing, there are about 60 countries participating in WorldWideScience.org, providing access to 50+databases and information portals. Not all content is in English. (Bronson, 2009)

Given this scenario, we focused on building a corpus constituted by the query logs registered by the *GreyGuide: Repository and Portal to Good Practices and Resources in Grey Literature*[1] *and received* by the ***WorldWideScience.org***[2] *(The Global Science Gateway) portal*: the aim is to retrieve information related to social media which as of today represent a considerable source of data more and more widely used for research ends.

This project includes eight months of query logs[3] registered between July 2017 and February 2018 for a total of 445,827 queries. The analysis mainly concentrates on the semantics of the queries received from the portal clients: it is a process of information retrieval from a rich digital catalogue whose language is dynamic, is evolving and follows – as well as reflects – the cultural changes of our modern society.

Methods and Tools

In order to analyze the available information a considerable pre-processing on four levels has been carried out:

- at the first level, the set of queries has been cleaned: duplicates, alphanumeric strings, strange graphical forms, IP addresses, etc. have been eliminated;
- at the second level, filters have been added and alphabetical order inserted for having a first picture of the contents of these queries;
- the third step consisted of several trials for choosing the focus;
- lastly, natural language processing (NLP) tools have been applied for processing the information and building the sample.

[1] http://greyguide.isti.cnr.it/ - *GreyGuide is the* online forum and repository of good practices and resources in Grey Literature. It was created - and is now edited - by GreyNet International (content provider) and ISTI-CNR, Pisa Italy (service provider): its launch was in December 2013 and since then *GreyGuide* provides a unique resource in the field of grey literature, which was long awaited and responds to the information needs of a diverse, international grey literature community. *Greynet International is one of the WorldWideScience Associate Members https://worldwidescience.org/alliancemembers.html.*

[2] https://worldwidescience.org/ - It is a global science gateway comprised of national and international scientific databases and portals. ***WorldWideScience.org accelerates*** scientific discovery and progress by providing one-stop searching of databases from around the world. WorldWideScience.org is maintained by the U.S. Department of Energy's Office of Scientific and Technical Information as the Operating Agent for the WorldWideScience Alliance.

[3] The General *Query Log* is the record of each SQL statement received from clients, in addition to their connection and disconnection time.

Since the corpus is made of queries collected in only eight months and the cleaning process reduced them consistently, as a result the final is relatively small. In addition, only the queries in English have been registered while those in other languages have been eliminated (there are a few in French, Spanish, Italian, Portuguese, Polish, Albanese, Galician, Corsican, and so on). As a curiosity, some queries also deal with socio-political or historical events (1915 Mexico Guerra, 1929 crisis unidos, 1960s economy, 1963 Sicily earthquake[4], 1979 Iran revolution, 1979 revolucion irani, 1984 George Orwell, 1986 FBI Shootout).
Coming to the NLP analysis, the software team has decided to follow these two steps:

1. free information extraction: it measured the frequency of all the words contained in the corpus. This preliminary investigation provided us with the whole scenario of the lexical variety of the queries and allowed us to focus on a set of terms from which we built a micro-ontology with meaningful terms relating to the queries launched on the portal;
2. ontology-based extraction: the extraction has been performed again using this micro-ontology which has been essentially used for enriching the domain. In this way, the search engine retrieved each single occurrence of those terms (monograms, bigrams, trigrams) which can be found starting from the ontology.

At the end of this pre-processing phase, we chose to focus on a flow of queries launched on the *WorldWideScience* platform concerning only the bigram *social media.*
Why social media?

- ✓ *nowadays social media are obviously a very effective means of communication but can even vehiculate knowledge as their various types (*eg.: blogs, YouTube*, Facebook, Twitter, etc.) are by now often quoted in bibliographical references amongst the more traditional categories (books, journals and so on);*
- ✓ *the subject involves document types pertaining to Grey Literature.*

An example of this type of extraction is given by some terms which belong to the fields of medicine and psychiatry and are paired with the bigram *social media*:

<***cyberbullying*** *social media*>
<***depression*** *social media*>
<***eating disorder*** *social media*>
<***negative effects*** *social media young adults*>
<***anxiety*** *social media*>
<*social media* **compulsive buying**>
<*social media* ***distraction***>
<***fake news*** *social media*>
<*social media* **millennial**>

At once, these terms show a negative connotation in relation to the use of social media: this phenomenon seems rather relevant in the queries of our corpus and therefore deserves further investigation.

Conclusion

The poster will illustrate the main linguistic features of *the Global Science Gateway by showing*:

- ✓ the lexical map representing the most used/recurrent words (in terms of occurrences) as well as the adoption of neologisms (a very interesting one is “netnography”) and hapaxes (such as “hastag”) in the realm of queries on social media;
- ✓ the comparison among the various typologies of social media on polarity[5], similarity and diversity.

[4] It is actually referring to the Sicily earthquake of 1693, not 1963.
[5] “In linguistics, a **polarity item** is a lexical item that can appear only in environments associated with a particular grammatical polarity – affirmative or negative. A polarity item that appears in affirmative (positive) contexts is called a **positive polarity item** (PPI), and one that appears in negative contexts is a **negative polarity item** (NPI)”, Wikipedia.

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 projects. Member of board at Institute for Computational Linguistics. She is author and co-author a number of publications dealing with Computational Linguistics, Computational Terminology and Grey Literature. 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

Stefania Biagioni graduated in Italian Language and Literature at the University of Pisa and specialized in Data Processing and DBMS. She is currently an associate 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 currently involved in the activities of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She has been head librarian of the Multidisciplinary Library of the CNR Campus in Pisa till August 2017. She was the responsible of ERCIM Technical Reference Digital Library (ETRDL) Project and currently is the coordinator 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 and grey literature. Her research interest are focused on digital libraries, knowledge sharing and transfer in scientific area, scholarly communication infrastructures, Open Access and Open Science. She has been dealing with grey literature since 90's. Since 2013 she is involved on the GreyGuide Project. Email: stefania.biagioni@isti.cnr.it
ORCID iD https://orcid.org/0000-0001-9518-0267

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
ORCID iD https://orcid.org/0000-0001-9808-6268

Visualizing Grey Literature Using a Bibliometric Approach

David E. Hubbard, Texas A&M University Libraries, USA

Grey literature finds its way into a topically diverse body of research and acquiring an overview of that body of literature – especially for the uninitiated - can be challenging, but visualization techniques can provide insights into what is a rather complex data set. This study seeks to explore how grey literature is discussed and used from topical perspective. Using journal and conference literature – bibliometric data – indexed in Web of Science and visualized using VOSviewer, a tool developed by the Centre for Science and Technology Studies at Leiden University for bibliometric analyses, an analysis of journal and conference literature mentioning grey literature within titles and abstracts was conducted. These topical visualizations, coupled with standard bibliometric approaches, can provide additional context and outline research trends and themes. Preliminary findings indicate an increased mentioning of grey literature in published research, especially in the biological and medical sciences. With respect to the medical sciences, much of the increase is due to the rise of systematic reviews since some of those studies often refer to searching the grey literature. There were far fewer articles and papers mentioning grey literature in the context of STEM disciplines, though "data" was mentioned in a variety of contexts. There are limitations to this approach in terms of the bibliometric data source, primarily published journal articles, rather than grey literature itself though the intent of this study was to see how grey literature was being mentioned within the more traditional scholarly literature. Another limitation is that such visualizations only provide a broad overview and are subject to interpretation, however, such analyses can point to areas that warrant further investigation.

Data Papers: Report on a Training Module Developed for Long Tail Research Data

Dominic Farace and Jerry Frantzen,
GreyNet International, Netherlands

In 2017, GreyNet embarked on a Data Papers Project in an effort to comply with the FAIR Principles endorsed by data repositories and archives worldwide. Compliance with these principles demonstrate that data/sets are findable, accessible, interoperable, and reusable.

To this end, a standardized template was designed and implemented as a tool in enabling authors/researchers to create and publish data papers within the GreyNet community as well as other communities of practice involved in the long tail of research data.

The results of the project formed the basis for a training module that has since been delivered in four workshop settings from March through October 2018. The conference poster focusses on the stepwise approach carried out by GreyNet and demonstrates how the data paper - a grey literature document type - has become a trusted tool in research and data sharing. This training module is offered both standalone or in collaboration with other open data initiatives.

SPECIAL PANEL SESSION

WORLDWIDE SCIENCE*Alliance* AND GLOBAL SCIENCE GATEWAY, 2008-2018

Panel Moderator: Brian Hitson, Director, Office of Scientific and Technical Information; U.S. DOE

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

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

Growth of WorldWideScience.org – 2009

Between inception of the prototype in 2007 and 2009, WorldWideScience.org grew from:

- 10 countries represented… to 56 countries
- 15 searchable databases/portals… to more than 50 databases

In 2009, China joined the WorldWideScience Alliance

Increasing the need for multilingual translations capabilities…

Multilingual WorldWideScience.org – 2010

In partnership with Microsoft Research, WorldWideScience.org introduced multilingual translations capabilities in June 2010

Breaking Down Language Barriers and Accelerating Scientific Discovery

Multilingual Translations

The world's first "one to many" and "many to one" multilingual translations tool in science

- Most automatic translations are limited to translating from a single language into another single language
- WorldWideScience.org partnering with Microsoft® Translator enables true multilingual functionality

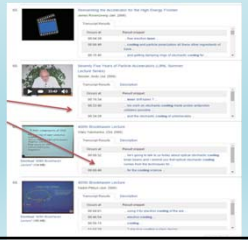
Dobrica Savić, Nuclear Information Section;
International Atomic Energy Agency, IAEA

Integration of Multimedia and Speech-Indexed Content – 2011

Multimedia (video, audio, images)

A major emerging form of scientific and technical information

Within WorldWideScience.org – users can search for the precise term within video and be directed to the exact point in the video where the term was spoken

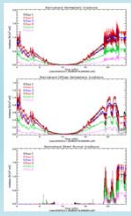


Integration of Scientific and Research Data – 2012

WorldWideScience.org overcomes many of the challenges associated with the discoverability of scientific and research data:

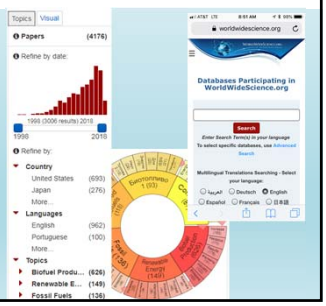
- Open Data sources searchable with a single query
- Consolidated results list
- Links to original records, including datasets

TTT F Phe TTC F Phe TTA L Leu TTG L Leu i TCT S Ser TCC S Ser TCA S Ser TCG S Ser TAT Y Tyr TAC Y Tyr TAA * Ter TAG * Ter TGT C Cys TGC C Cys TGA * Ter TGG W Trp CTT L Leu CTC L Leu CTA L Leu CTG L Leu i CCT P Pro CCC P Pro CCA P Pro CCG P Pro CAT H His CAC H His CAA Q Gln CAG Q Gln CGT R Arg CGC R Arg CGA R



Growth and Enhancements – 2013-2018

- **New design and layout, including mobile responsive compatibility**
- **New databases and portals added – over 100 resources now searchable**
- **Federated search technology improvements – faster search speeds and addition of results clustering options**



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

Growth and Enhancements – 2013-2018

Inclusion of Open Access/Public Access Resources

For the past 10 Years…

WORLDWIDESCIENCE.ORG
The Global Science Gateway

WorldWideScience.org has accelerated scientific discovery by:

- Providing access to authoritative, high quality resources
- Enabling users to search multiple databases with a single query, including unique deep web content not readily indexed by major search engines
- Translating search queries into multiple languages, and then translating consolidated results into the user's native language
- Supporting discovery of textual and non-textual (multimedia, research data, scientific software) information

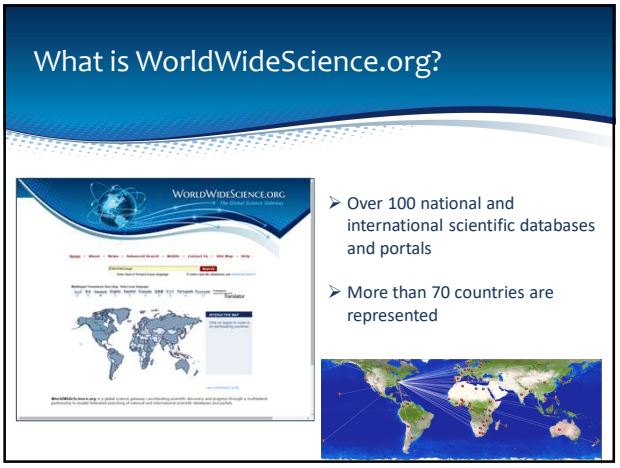
The *Next* 10 Years…

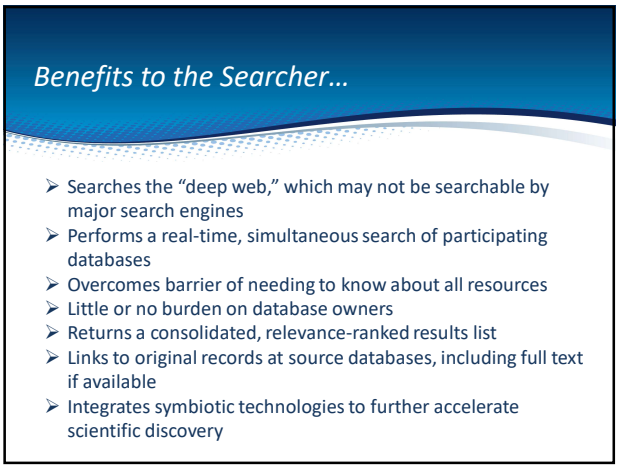
WORLDWIDESCIENCE.ORG
The Global Science Gateway

WorldWideScience.org will promote scientific collaboration, participation, and transparency:

- Build on the base: *More* sources, countries, languages, full text, multimedia, data, and software
- Provide equal access to science for anyone on the Internet through continued growth of open/public access resources for scholarly publications, research data, and scientific software
- Multilingual federated search of these diverse and related research objects continues to have major discovery advantages over traditional search engines
- Increase membership and participation, both as database/source owners and as WorldWideScience Alliance members

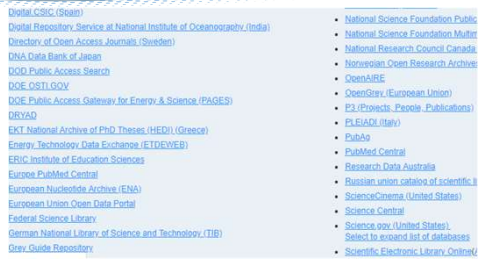
Lorrie Johnson, Office of Scientific and Technical Information;
U.S. Department of Energy (DOE)





Lorrie Johnson, Office of Scientific and Technical Information;
U.S. Department of Energy (DOE)

A variety of databases and formats…

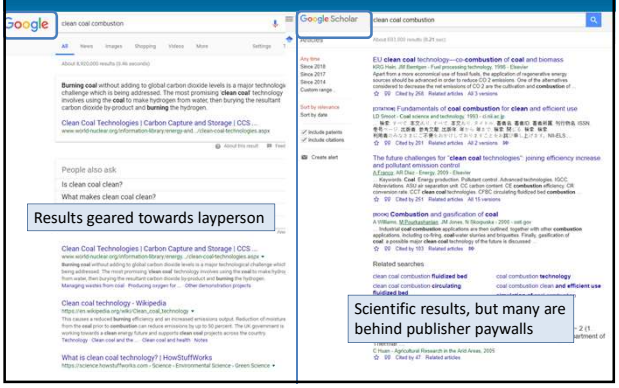


- Over 60 English language databases from around the world
- 20 Multilingual (non-English) databases
- Multimedia sources – audio and visual content
- 15 Research Data Sources and over a dozen Public Access resources
- ***And, most recently, scientific SOFTWARE and CODE***

A Measure of WorldWideScience.org's Uniqueness

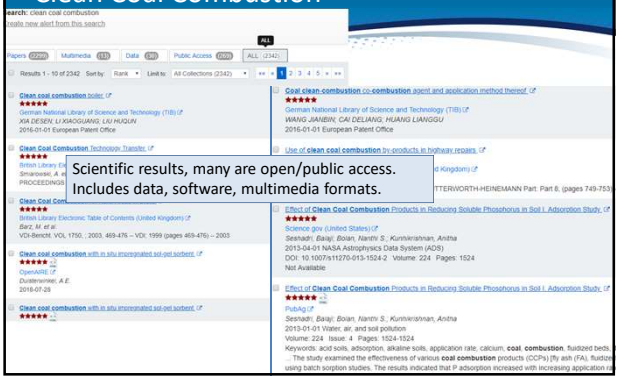
- 33 sample queries launched in Google, Google Scholar, and WorldWideScience.org
- Similar quantities in the numbers of results, but ***very little overlap***
- Among the "top 50" results from each search engine, only ~10% overlap – or **90% uniqueness – in WorldWideScience.org results**

Google and Google Scholar Results for: Clean Coal Combustion



Lorrie Johnson, Office of Scientific and Technical Information;
U.S. Department of Energy (DOE)

WorldWideScience.org Results for: Clean Coal Combustion



WorldWideScience.org Topic Pages

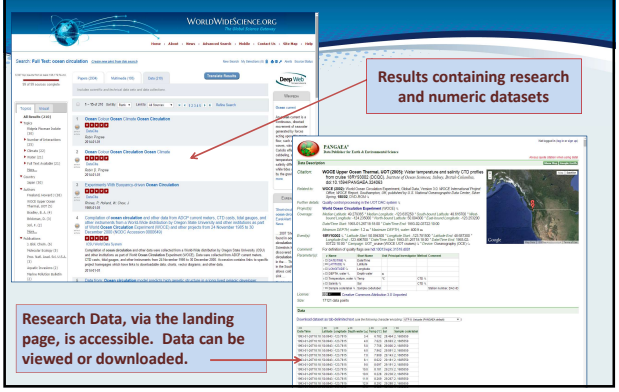
Utilizing the power of Google, Bing, Yahoo, and other search engines:

- Millions of Topic Pages created
- Exposes distributed WorldWideScience.org content to commercial search engines via site map protocols

Multilingual Translations: Query is translated into languages of all databases, searches are completed, and results are combined into a single, relevance-ranked results list. User translates results into their preferred language.

Lorrie Johnson, Office of Scientific and Technical Information;
U.S. Department of Energy (DOE)

Integration of Scientific Research Data

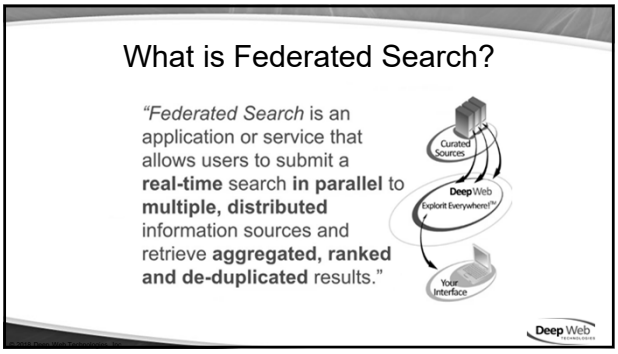


WorldWideScience Supports Open Science

User Friendly Features:

- Basic and Advanced Search Options
- Ability to search selected databases (via Advanced Search)
- Results tabs for Text, Multimedia, Data/Software, Public Access
- Sort by Rank, Date, Title, Author
- Limit results to specific databases
- Refine results by Topic, Authors, Country, Document Type (full text), Language, etc.
- Textual and visual clustering capabilities
- Mark/Save records
- Routine, customizable Alerts service

Abe Lederman, Deep Web Technologies, Inc.



Abe Lederman, Deep Web Technologies, Inc.

Anatomy of a Search

1. User fills out search form for simple or advanced search
2. WWS searches selected sources in parallel
3. Results are aggregated from selected sources
4. Results are ranked
5. Results are displayed incrementally, as they are received
6. Users can narrow and filter results
7. Alerts monitor sources for relevant new results
8. Multilingual search in 10 languages handles query and search title and summary translation

Parallel Searching and Incremental Results

- Explorit queries multiple sources simultaneously
- Results are merged together (aggregated) as they are received from the sources
- Results are then ranked for relevance and displayed
- Our asynchronous API enables the display of incremental results

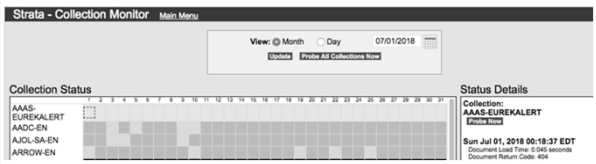
Smart Searching

Connectors are the engine behind searching. They …

- Are custom crafted for each source
- Translate Explorit! query syntax into source's query syntax
- Extract metadata returned in the result list into Explorit! display fields and cluster fields
- Retrieve multiple pages of results from a source

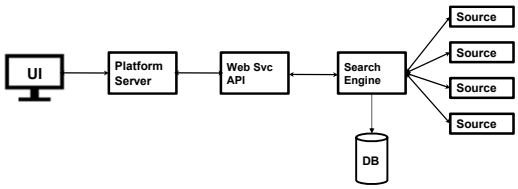
Abe Lederman, Deep Web Technologies, Inc.

Connector Monitoring



- Perform a set of tests against each connector several times a day
- Alert the Connector Group of failures detected

Building Blocks

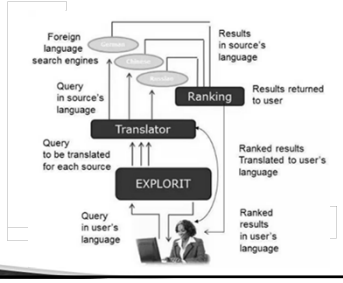


Ranking: the Secret Sauce

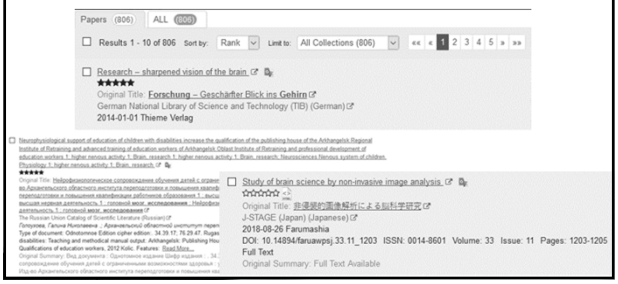
- Creating Root-Words: Stemming
- Conducting Relevance Weighing
  - Position, Density, Proximity, Same Order
  - Boost rank of some sources
  - Boost rank of more current results
- Proprietary Algorithms

Abe Lederman, Deep Web Technologies, Inc.

Translation Flow



Translated Results



Federation of Federation

- Explorit can federate other federated search applications
- Federation of federation is a way to divide and conquer and potentially search thousands of sources
- WWS searches Science.gov and all of its sources

Deep Web

Justin Fessler, IBM Corporation, United States

IBM WATSON

IBM Watson

Justin Fessler – Artificial Intelligence Leader, IBM Federal

1 © 2018 IBM

IBM WATSON

3 top trends/challenges in AI and what's coming/available today…

(AI: Artificial Intelligence or Augmented Intelligence….?)

2 © 2018 IBM

IBM WATSON

Challenge #1

Information Access

Employees struggle to get a complete, secure view of all relevant data and information (structured/unstructured) whether internal, external, OSINT, etc.

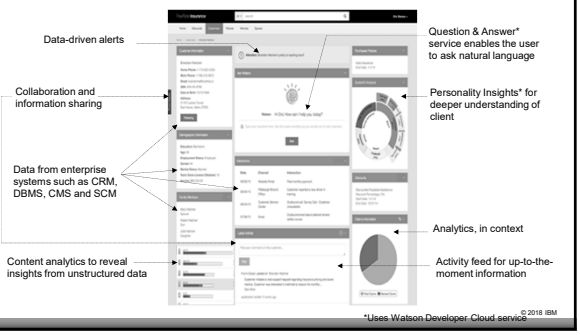
3 © 2018 IBM

Justin Fessler, IBM Corporation, United States

IBM WATSON

360-degree information application

Information, analytics and cognitive insights presented in context



*Uses Watson Developer Cloud service

4 © 2018 IBM

IBM WATSON

Challenge #2

Unstructured Content

80% of data is unstructured but only a small percentage leveraged for insights in uncovering patterns within investigations, inspections, condition-based maintenance, etc.

5 © 2018 IBM

Solution #2

IBM WATSON

Smart is: breakthrough content analysis

IBM Watson (Jeopardy)

" … an information seeking tool that ' s capable of understanding your question to make sure you get what you want and then deliver ' s that content through a naturally flowing dialog "

Dr. David Ferrucci
Principal Investigator
Watson project

Business Challenge
Advance the state of the art in broad domain Question Answer (QA) systems to enable breakthrough applications in many different industries.

What's Smart?
Uses IBM Content Analytics (LanguageWare) in conjunction with other technologies to read, analyze and understand vast sources of unstructured content. Runs many algorithms in parallel to create, compare and determine confidence in candidate answers. Presents answers with a confidence level attached.

Smarter Business Outcomes
Coming to your industry soon! Will deliver value in limitless applications starting with clinical healthcare, customer care, government intelligence and beyond.

Industry context: ***broad industry value***
Value driver: ***improve business decisions***
Solution onramp: ***content analytics***

6 © 2018 IBM

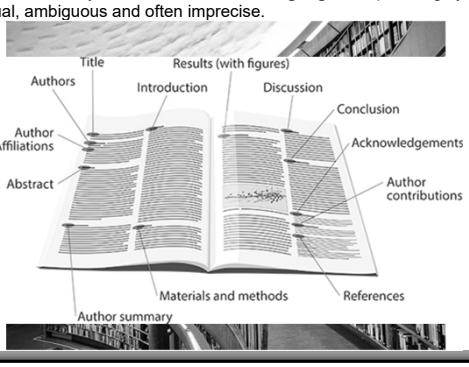
Justin Fessler, IBM Corporation, United States

Solution #2

IBM WATSON

What Computers Find Hard

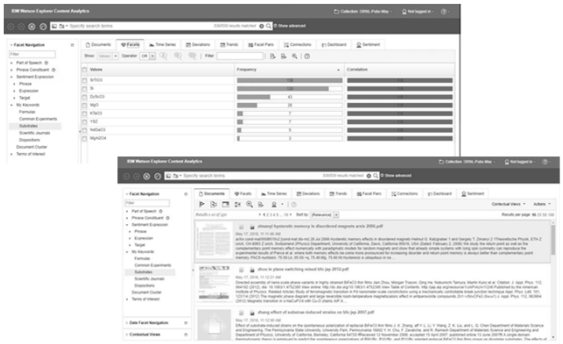
Computer programs are natively **explicit, fast** and **exacting** in their calculation over numbers and symbols….But **Natural Language** is implicit, highly contextual, ambiguous and often imprecise.



7 © 2018 IBM

Solution #2

IBM WATSON



***Materials science journals provided by Oak Ridge National Lab

8 © 2018 IBM

IBM WATSON

Challenge #3

Scaling expertise through training and Predictive Analytics

Continuous pressure to increase performance and innovation—while doing more with less

9 © 2018 IBM

Justin Fessler, IBM Corporation, United States

IBM WATSON

Unstructured data is messy but filled with key facts

PC 143 (Hunter)
15 June 2006 23:47
Suspect identified himself as **John Setsuko**. Matched description given by night club doorman (IC1, Male, Ag 22-24 yrs, blue Everton shirt). Stopped whilst driving **White Ford Mondeo**, **W563 WDL**. **Address** given as **22 East Dene Ridge, Copdock, Ipswich**. Searched at scene and found in possession of **1oz Cannabis Resin** and **lockable pocket knife**.

10 © 2018 IBM

IBM WATSON

Historic Data

$$\begin{aligned}\mathbb{P}(S_n - \mathrm{E}[S_n] \geq t) &= \mathbb{P}\left(e^{s(S_n - \mathrm{E}[S_n])} \geq e^{st}\right) \\ &\leq e^{-st}\mathrm{E}\left[e^{s(S_n - \mathrm{E}[S_n])}\right] \\ &= e^{-st}\prod_{i=1}^{n}\mathrm{E}\left[e^{s(X_i - \mathrm{E}[X_i])}\right] \\ &\leq e^{-st}\prod_{i=1}^{n} e^{\frac{s^2(b_i-a_i)^2}{8}} \\ &= \exp\left(-st + \tfrac{1}{8}s^2\sum_{i=1}^{n}(b_i - a_i)^2\right)\end{aligned}$$

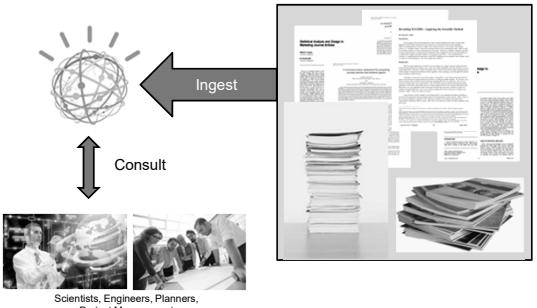
lictions

11 © 2018 IBM

Use case #1

IBM

Watson Aerospace Innovation Advisor – NASA Langley



© 2016 IBM Corporation

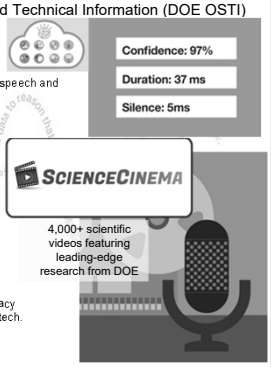
Justin Fessler, IBM Corporation, United States

Use case #2 IBM WATSON

US Department of Energy, Office of Science and Technical Information (DOE OSTI)

Audio Indexing

- **What is Audio Indexing?**
  - Processing of audiovisual data that extracts text from speech and makes it searchable
- **IBM Cloud:** Watson Bluemix
  - WAV, WMA, MP3, AAC, WMV, MP4
  - In the process of converting audio and video formats to MP3 and MP4
- **Features**
  - Search for specific words or phrases
  - System returns audio snippets
  - Viewer can go to the proper place in the video
  - Captions are provided throughout each video
- **Future Goals:**
  - Increased speed and accuracy of processing
  - Custom Vocabulary creation to further increase accuracy
  - Natural Language Processing features using Watson tech.
  - Language options– French, Spanish, etc.



13 © 2018 IBM

IBM WATSON

THANK YOU

Justin Fessler
NA Leader, Data Science
US Federal, Comms/CSI, Commercial, E&U, Canada
703-231-5326
jfessle@us.ibm.com

14 © 2018 IBM

A Case Study on the Content Curation for Improving Effectiveness of Research Reports

Seokjong Lim and Heykyong Hwang
Korea Institute of Science & Technology Information, KISTI, South Korea

The Korean government has named Research Outcome Management & Distribution Agency in accordance with 'Rules on Management of National R&D Programs' in 2008 for the efficiency of management and use of research outcomes. There is a growing importance on the role of agency of National R&D information management and sharing, and nationwide management of the performance of national R&D outcomes. So far the national R&D reports have been managed over the past 10 years in Korea. To enhance an utility of the content within the research reports, we developed an actual case on the content curation by identifying content object such as author name, institution name, subject, reference literature, digital object, funding data. This implied a new paradigm of management of R&D reports as a grey literature. Most of these content objects provide a linkage to original reports and other digital objects which has been identified by DOI. And also, several works on the non-text object within the original reports have attempted to find a detailed object like tables and graphics. It was something difficult to find object within existing reports and have only been used on the level of metadata and full-text of them.

We show that the visualization of summary graphs using keywords, the data set using curated content, and the evolution of data acquisition and management through the content curation.

The results point to the effective content sharing and use on the National R&D outcomes. This case study would be an advanced curation model in the field of grey literature management.

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

Open Data engages Citation and Reuse: A Follow-up Study on Enhanced Publication

Dominic Farace, Jerry Frantzen, GreyNet International, Netherlands;
Joachim Schöpfel, University of Lille, France

In 2011, GreyNet embarked on an Enhanced Publications Project (EPP) in order to link its collection of full text conference papers with accompanying research data. The initial phase in the study dealt with the design and implementation of an online questionnaire among authors, who were published in the International Conference Series on Grey Literature. From 2012 onwards, subsequent phases in the project dealt with the acquisition, submission, indexing, and archiving of GreyNet's collection of published datasets now housed in the DANS EASY data archive.

In 2017, GreyNet's Enhanced Publications Project was further broadened to include a Data Papers Project. Here, emphasis focused on describing the data rather than analyzing it. As such, the data paper signals data sharing and in this way promotes both data citation and the potential reuse of research data in line with the FAIR Guiding Principles for scientific data management and stewardship.

Available results from the Data Papers Project presented last year at GL19 concludes where this study commences. Here, we now seek to demonstrate the reuse of survey data collected in 2011 combined with survey data that will be newly collected via an online questionnaire. The survey population will be drawn from among GreyNet's author base; and, a selection of questions from the 2011 Survey will be joined by newly formulated questions in constructing the questionnaire. Furthermore, GreyNet relying upon available use and usage statistics compiled from various sources will seek to provide evidence of data citation and referencing. The results of this study are expected to demonstrate an increased willingness among GreyNet authors to share their research data – this in part due to GreyNet's program of enhanced publication embedded in its workflow over the past six years. The study will provide an example of the reuse and further comparison of the results of survey data, which can be incorporated in GreyNet's program of training and instruction. However, statistics on data citation and referencing are less likely expected to provide indicative results.

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 (1993-2018). 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 **ORCID iD** https://orcid.org/0000-0003-2561-3631

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 **ORCID iD** https://orcid.org/0000-0002-3405-7078

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 **ORCID iD** https://orcid.org/0000-0002-4000-807X

Open Data engages Citation and Reuse:
A Follow-up Study on Enhanced Publication

Dominic Farace and Jerry Frantzen
GreyNet International, Netherlands

Joachim Schöpfel
University of Lille, France

EPP

Background
Enhanced Publications Project
2011

Enhanced publication inherently contributes 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.

1. Design of the Questionnaire and Author Survey
2. Acquisition and Submission of Research Data
3. Data upload and cross-linking OpenGrey ↔ DANS Archive
4. Draft of EPP guidelines and the design of a workflow

Follow-up
Data Papers Project
2017

Findable Accessible Interoperable Reusable

DATA papers

Data Papers

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).

Author Survey on Research Data
2018

We sought to demonstrate the reuse of survey data collected in 2011 combined with survey data that was newly collected via an online questionnaire.

A selection of questions from the 2011 Survey was joined by newly formulated questions in constructing the questionnaire.

Survey Population
2011 and 2018

The survey populations were drawn from among GreyNet's author base:

Survey Population	First Authors	Survey Recipients	Survey Respondents	Survey Results %
2011	162	95	50	52,6%
2018	115	94	44	46,8%

Questionnaire
2018

The 2018 questionnaire consisted of 10 questions, 6 of which appeared on the 2011 questionnaire. The reused questions dealt with the author's own empirical research data, its availability, the formats in which it appears, and the author's willingness to archive it and make it openly accessible.

The 4 additional questions dealt with the respondent's citation and reference to data, their use of data journals in carrying out search and retrieval, and whether they (co)authored a data paper or data article.

Survey Results
2011 - 2018

Does one or more of your conference papers in the GL-Series base its findings on empirical or statistical data?

2011	Question 1	2018	Question 1
Yes	60%	Yes	70%
No	40%	No	30%
	100%		100%

Data Analysis
2011-2018

Significant Changes

- More data and/or datasets are available in part or whole for archiving purposes (54% → 70%) p = .005
- More authors prefer entering their data and/or datasets directly in DANS (18% → 42%) p = .05
- The data and data formats are more specific (44% → 76%) p = .1

Data Analysis
2011-2018

No Changes or Insignificant Changes

- 50% of the respondents are willing to submit data, datasets, or subsets to DANS
- 40% of the respondents prefer that GreyNet enter their data and/or datasets in DANS

Conclusions

- Not all papers are based on empirical data
- Not all data can be shared
- Some authors have institutional practice
- Growing data awareness and literacy
- A community in transition

Lagniappe

Open Data engages Citation and Reuse

- Research Data published prior to the Research Paper
- Data Paper published prior to the Research Paper
- ORCID and DOI persistent identifiers further included in the GL-Conference Papers
- Results of this Follow-up Study included in the GreyForum Training and Workshops

Thanks!

'Now Let's Get Your Data Cited'

info@greynet.org

The Q-Codes: Metadata, Research data, and Desiderata, Oh My! Improving Access to Grey Literature in Family Medicine

Melissa P. Resnick, University of Texas Health Science Center at Houston, Houston, TX, USA
Ashwin Ittoo, HEC Management School, University of Liège, Belgium
Marc Jamoulle and Marc Vanmeerbeek, Department of General Practice, University of Liège, Belgium
Frank S. Shamenek, Consultant, New York, USA
Chiehwen Ed Hsu, College of Management, National CK University, Tainan, Taiwan
Robert Vander Stichele, Heymans Institute of Pharmacology, University of Ghent, Belgium
Julien Grosjean and Stefan Darmoni, Department of Information and Medical Informatics (D2IM), University of Rouen, France
Elena Cardillo, Institute of Informatics and Telematics, National Research Council, Italy
Miguel Pizzanelli, Department of Family and Community Medicine, University of the Republic of Uruguay

Problem/Goal: In GL19's "Indexing grey multilingual literature in General Practice in the era of Semantic Web," Jamoulle and colleagues propose the use of a relatively new terminology (3CGP) to allow for the indexing and retrieval of (GP/FM) knowledge which otherwise would be lost, or difficult to locate. Though designed to meet Cimino's (1998) twelve desiderata for the design of a controlled healthcare vocabulary, Jamoulle and colleagues acknowledge that a detailed requirement by requirement evaluation of 3CGP was not performed. The goal of this paper is to evaluate the Q-Code component of the 3CGP terminology, in detail, with each of Cimino's twelve desiderata.
Research Method/Procedure: In our work, we will focus on qualitative analysis, whereby our taxonomy, the Q-Codes, and in particular, its vocabulary satisfies a standard set of desiderata. Qualitative analysis provides a simple and yet effective way to assess our taxonomy's quality. We will briefly describe each of the desiderata and discuss how our taxonomy satisfies each one of them (or not).
Anticipated Results of the Research: The qualitative evaluation is intended as an initial stage, which focuses on our taxonomy's contents, namely, its vocabulary (e.g. terms and definitions). Our aim with the qualitative evaluation is to investigate whether our proposed taxonomy, and in particular its vocabulary, satisfies a set of desiderata. This will enable us to determine whether the knowledge acquisition and (part of) the conceptualization steps of our ontology development process have been performed correctly. We consider that validating our vocabulary against a set of well-defined desiderata is paramount before evaluating other aspects of the taxonomy (such as the relations). As a set of desiderata, we chose that proposed by Cimino in his seminal study entitled "Desiderata for controlled medical vocabularies in the twenty-first century" (Cimino, 1998). These desiderata ensure that our taxonomy can be successfully deployed and exploited in actual GM/FM applications / activities, such as indexing grey literature. The desiderata define a set of (desired) characteristics that (ideally all) standard medical vocabularies should satisfy. Thus, these desiderata help in alleviating inter-operability issues, with the use of common standards ensuring the efficient integration of our taxonomy with other medical vocabularies and resources (taxonomies, ontologies). From the results of this study, improvements can be made to the Q-Codes component of, and thus, the 3CGP terminology. This, in turn, improves the ability to index the grey literature with the 3CGP terminology, providing greater access to needed information.
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)

Analysis of folk literature in grey literature from the National Library of China

Cui Yue, National Library of China, Beijing, China

As a nationwide library, national bibliographic center and document information center, the National Library of China (NLC) is not only comprehensively collecting official Chinese publications but has also, long before, attached importance to the acquisition of grey literature.

The NLC's acquisition of grey literature has changed from a basic, scattered collection to a comprehensive collection and finally to a key collection. The institution was established through, in succession, the academic dissertation library, local chronicles and the genealogical documents collection.

At the same time, the NLC was exploring the construction of characteristic Chinese grey literature. Through the analysis of the present grey literature collections in NLC, the amount of folk literature, conference proceedings, research reports and document assembly has already taken shape.

Therefore, a new grey literature system has been formed, among which the most significant increase is in folk literature with distinctive regional characteristics and historical value.

Broadly speaking, folk literature refers to all literature bearing historical and cultural information and kept in folkloric. It contains the literature produced by the people and the official literature lost in folkloric.

This literature covers historical, cultural, artistic and other fields, such as folk poetry, intangible cultural heritage, memoirs, revolutionary historical materials, etc.

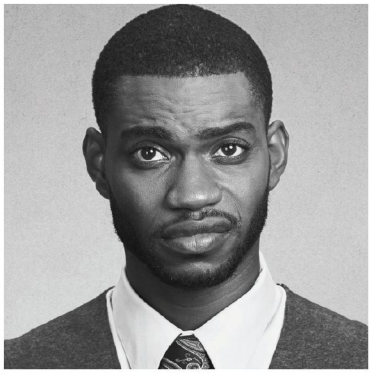
At present, literature collections and sorting institutions, such as libraries, are not collecting and researching of folk literature to a great enough extent. As a result, researchers cannot easily find comprehensive literature. As a library that bears a mission of preserving cultural heritage, the collection and utilization of folk literature should be emphasized.

Taking folk literature in the grey literature collected by NLC over the past 10 years as a sample, this paper analyzes the content, composition, quantity, type, geographical distribution, value and significance of this kind of literature. To solve the problem of the construction of folk literature in grey literature, this paper innovates acquisition methods and the collection scope of folk literature within grey literature. Some ideas and suggestions are put forward on how to establish folk literature collection institutions and rationally utilize and develop existing resources.

PsycEXTRA®

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

www.ebsco.com ▪ (978) 356-6500 ▪ (800) 653-2726 ▪ information@ebsco.com

When the Virtual Becomes Reality: An Environmental Scan of the Presence of Virtual Reality and Artificial Intelligence in Health and Cancer Care Environments

Marcus Vaska, Alberta Health Services; Knowledge Resource Service, Canada

Grey literature has long been associated with technological enhancements, recognizing the power that informational communication, namely, social media, plays in generating interest in blogs, Twitter feeds, and other instantaneous knowledge exchange platforms. The ability of these programs to generate and identify specific data patterns from a single posting has led to increasing interest in two aspects of machine learning in health care, namely Artificial Intelligence (AI) and Virtual Reality (VR). AI "mimics elements of human cognition by computational means", whereas VR enhances this cognition by allowing users to interact with a "three-dimensional, computer generated environment", manipulating objects and scenarios in an artificial world. Introduced as a form of grey literature via Second Life, a popular role-playing online world launched in 2003, VR and AI have had a visible presence in numerous sectors, including healthcare.

In 2011, IBM created *Watson,* a supercomputer considered to be one of the most revolutionary breakthrough's in artificial intelligence. To test this claim, *Watson* appeared on an episode of *Jeopardy*, one of the longest-running game shows in the United States, in a friendly competition match between two of the winningest contestants in the show's 50 year history. *Watson's* emphatic victory over the human contestants drew increasing interest to other applications of artificial intelligence and virtual reality, specifically in the field of healthcare.

While the first use of AI and VR in medicine is believed to have occurred in the 1990s for interpreting electrocardiograms, the invention of cloud networking in 2006 is considered the first proven use of AI and VR in the modern era focusing on healthcare. Although the arguments for AI and VI in clinical settings are plentiful, ranging from enhancing imaging and increased processing speed in electronic medical record (EMR) applications, the scenario is less clear-cut within the environment of cancer care. At the 2016 International Symposium of Biomedical Imaging in Prague, a joint team of scientists and engineers claimed that the use of artificial intelligence resulted in a "92% accuracy [rate of detection] in breast tissue cancer cells." However, a column authored in 2017 disputed a claim by IBM that *Watson* was the new revolution to cancer care.

This paper will aim to shed light on how artificial intelligence and virtual reality is viewed in both health and cancer care fields via a two-fold environmental scan approach, namely an anonymous survey polling staff working at two cancer care facilities in Calgary, Alberta, Canada, asking respondents to comment on any papers they have ever encountered in their own practice/research discussing AI or VR. This practice will be supplemented with a comprehensive search through the academic literature to achieve a hoped-for grand total of 50 unique papers. Each of these papers will be analyzed via the use of Altmetrics, "a single research output [that] can be talked about across dozens of different platforms", a methodology introduced by Schöpfel and Prost at GL18, to determine how these perceived core papers are being shared via the use of social media.

Bionote

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 **ORCID iD** https://orcid.org/0000-0002-4753-3213

A department of System Innovations and Programs

When the Virtual Becomes Reality: An Environmental Scan of the Presence of Virtual Reality and Artificial Intelligence in Health and Cancer Care Environments

Marcus Vaska (marcus.vaska@ahs.ca)

GL20, December 3-4, 2018

© Scanrail 123RF.com

Definitions: Artificial Intelligence & Virtual Reality

- **Artificial Intelligence**: "mimics elements of human cognition by computational means" [Rubak 2018]
- **Virtual Reality**: "a three-dimensional, computer-generated environment, which can be explored or interacted with by a person" [Virtual Reality Society 2017]

© olegdudko 123RF.com

AI in Healthcare: A Brief History

- Tautology: "collaborative human-machine tasking improves performance over either alone" [Miller and Brown 2018]
- 1990s: first use of AI in medicine (electrocardiograms)
- The diagnostic speed and accuracy of AI in medical images parallels that of medical experts [BENEFIT]
- AI uses natural language processing when reading electronic medical records, being able to note any errors made by human bias[BENEFIT]
- "AI is neither astute nor intuitive [thus ensuring] physicians will remain essential to cognitive medical practice" [Miller and Brown 2018] [CHALLENGE]

VR in Cancer Care: A Brief Retrospective

- 1997: proposal launched to develop a more effective and less invasive colonoscopy [Wells, 2016]
- 2011: clinical trial conducted at the M.D. Anderson Cancer Center, University of Texas→virtual reality intervention in cancer genetics [M.D. Anderson Cancer Centre, 2011]
- 2016: 3-D virtual reality colonoscopy pioneered at the University of California, San Francisco [Wells, 2016]
- 2017: VR introduced in the Adolescent and Young Adult (AYA) cancer program at the University of South Carolina (USC) Norris Comprehensive Cancer Center [Hu, 2017]

Two Environmental Scans

One Minute Question: Please provide the citations of any papers that you have ever encountered in your practice/research which discuss the use of virtual reality or artificial intelligence (either specifically in cancer or more generally in public health)

AI & VR in Cancer Care OR Health Care: Literature Search

Resources Consulted:

- MEDLINE (Ovid)
- PubMed
- CINAHL
- MEDLINE (Ebsco)
- Google Scholar

Limits Applied:

- English language
- 2007 – present

Terms Brainstormed:

artificial intelligence
virtual reality
machine learning
healthcare
public health
cancer
oncology
technology
2-D; 3-D

Environmental Scan Results

- Oncologists in the Calgary Zone
  - **12/110** responded, suggested **13** paper citations
- Literature Search on AI & VR in Cancer and Health
  - **39** papers identified
- Total Papers for Analysis: **52**
- Representative Sample→Altmetric Scores
  - **5** papers pertaining to **Cancer Care**
  - **5** papers pertaining to **Healthcare**
- Capture grey literature document types via social media:
  - Blogs
  - Facebook pages
  - News Outlets
  - Twitter feeds

AI in Healthcare: 5 Selected Papers for Analysis

- Garret, B., Taverner, T., Masinde, W., Gromala, D., Shaw, C., & Negraeff, M. (2014). A rapid evidence assessment of immersive virtual reality as an adjunct therapy in acute pain management in clinical practice. *Clinical Journal of Pain, 30*(12): 1089-1098.
- Keller, M., Park, H., Cunningham, M., Fouladian, J., Chen, M., & Spiegel, B. (2017). Public perceptions regarding use of virtual reality in health care: a social media content analysis using Facebook. *Journal of Medical Internet Research, 19*(12): e419.
- Miller, D., & Brown, E. (2017). Artificial intelligence in medical practice: the question to the Answer? *American Journal of Medicine, 131*(2): 129-133.
- Wiederhold, B., Gao, K., Sulea, C., & Wiederhold, M. (2014). Virtual reality as a distraction technique in chronic pain patients. *CyberPsychology, Behavior, & Social Networking, 17*(6): 346-352.
- Wiederhold, B., Riva, G., & Gutierrez-Maldonado, J. (2016). Virtual reality in the assessment and treatment of weight-related disorders. *CyberPsychology, Behavior, & Social Networking, 19*(2): 67-73.

KRS@ahs.ca
KRS.albertahealthservices.ca

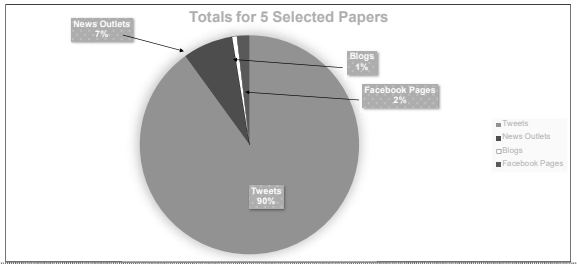
Alberta Health Services
Knowledge Management

AI in Healthcare: Analysis & Discussion

KRS@ahs.ca
KRS.albertahealthservices.ca

Alberta Health Services
Knowledge Management

AI in Healthcare – Grey Literature Document Types



KRS@ahs.ca
KRS.albertahealthservices.ca

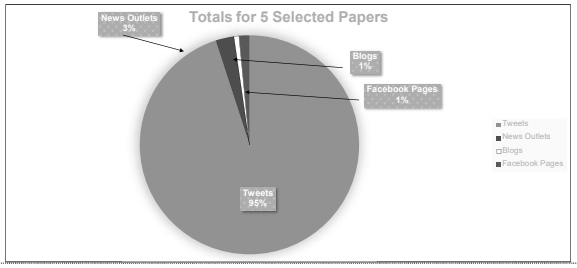
Alberta Health Services
Knowledge Management

VR in Cancer Care: 5 Selected Papers for Analysis

- Chirico. A., Lucidi, F., De Laurentis, M., Milanese, C., Napoli, A., & Giordano, A. (2015).Virtual reality in health system: beyond entertainment. A mini-review on the efficacy of VR during cancer treatment. *Journal of Cellular Physiology, 231*(2): 275-287.
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L., & Aerts, H. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer* [Epub ahead of print]
- Li, W., Chung, J., & Ho, E. (2011). The effectiveness of therapeutic play, using virtual reality computer games, in promoting the psychological well-being of children hospitalized with cancer. *Journal of Clinical Nursing, 20*(15-16): 2135-2143.
- Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D., Barnholtz-Sloan, J., Velazquez, J., Brat, D., & Cooper, L. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences of the United States of America, 115*(13): E2970-E2979.
- Schneider, S., & Hood, L. (2007). Virtual reality: a distraction intervention for chemotherapy. *Oncology Nursing Forum, 34*(1): 39-46.

VR in Cancer Care: Analysis & Discussion

VR in Cancer Care – Grey Literature Document Types



Pre-Conference Announcement

Twenty-First International Conference on Grey Literature

Open Science Encompasses New Forms of Grey Literature

German National Library of Science and Technology

Hannover, Germany ● October 22-23, 2019

TIB LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY

List of Participating Organizations

As of November 8, 2018

Access Innovations, Inc.	United States
Air Force Civil Engineer Center AGEISS Inc.	United States
Alberta Health Services	Canada
Data Archiving and Networked Services, DANS-KNAW	Netherlands
Deep Web Technologies	United States
Duke University Medical Center Library	United States
EBSCO	United States
Federal Library Information Network; Library of Congress, FEDLINK-LOC	United States
Gambling Research Exchange Ontario, GREO	Canada
George Mason University, GMU	United States
Georgetown University Medical Center; Dahlgren Memorial Library	United States
German National Library of Science and Technology, TIB	Germany
GreyNet International, Grey Literature Network Service	Netherlands
IBM Federal	United States
Indiana University Libraries	United States
Information Today Inc.	United States
Institut de l'Information Scientifique et Technique, Inist-CNRS	France
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
Japan Atomic Energy Agency, JAEA	Japan
Korea Institute of Science & Technology Information, KISTI	Korea
LibSource, LAC-Group Company	United States
Louisiana State University Libraries, LSU	United States
Loyola University New Orleans	United States
National CK University; College of Management	Taiwan
National Institute of Standards and Technology, NIST	United States
National Library of China, NLC	China
National Library of Technology, NTK	Czech Republic
National Research Council of Italy, CNR Central Library	Italy
NeMIS Research Laboratory, ISTI-CNR	Italy
Network and Informative Systems Office, CNR	Italy
Nuclear Information Section; International Atomic Energy Agency, NIS-IAEA	United Nations
Office of Scientific and Technical Information; OSTI-DOE	United States
PricewaterhouseCoopers, PwC	Netherlands
Princeton University	United States
Sandia National Laboratories	United States
Slovak Centre of Scientific and Technical Information, CVTISR	Slovakia
Texas A&M University Libraries	United States
TextRelease, Program and Conference Bureau	Netherlands
University of California, Irvine Libraries	United States
University of Florida; George A. Smathers Libraries	United States
University of Ghent, Heymans Institute of Pharmacology	Belgium
University of Houston, University Libraries	United States
University of Illinois at Urbana–Champaign	United States
University of Liège, Department of General Practice	Belgium
University of Liège, HEC School of Management	Belgium
University of Lille	France
University of Maryland; University Libraries	United States
University of Minnesota Libraries	United States
University of Rouen, Department of Information and Medical Informatics	France
University of Southern California, USC	United States
University of Texas Health Science Center at Houston, School of Biomedical Informatics	United States
University of the Republic of Uruguay	Uruguay
University of Wisconsin, Milwaukee	United States
U.S. Geological Survey, USGS	United States
U.S. Government Accountability Office, GAO	United States
U.S. Government Publishing Office, GPO	United States
Utah State University, USU	United States
WorldWide Science Alliance	United States

the grey Journal

15 YEARS

International Journal on Grey Literature

Subscription Order Form

For *GreyNet* Members free or reduced rates, see http://www.greynet.org/membership.html

TGJ Volume 15, 2019	*Type of Subscription:*	*Amount in Euros*	*Total*
THE GREY JOURNAL - Printed ISSN 1574-1796 *Annual Subscription, including Postage and Handling*	☐ Institutional	€ 240	**€**
THE GREY JOURNAL - PDF/ CD-Rom ISSN 1574-9320 *Annual Subscription, including Postage and Handling*	☐ Institutional	€ 240	**€**
THE GREY JOURNAL - PDF/ Email / PWP ISSN 1574-180X *Annual Subscription, including Electronic Handling*	☐ Institutional	€ 240	**€**

Customer information

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

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

❑ Direct transfer to TextRelease, Rabobank Amsterdam, Netherlands
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to 'TGJ/The Grey Journal'

❑ 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 Ingenico|Ogone, a designated payment service for Visa Card and MasterCard.**

Correspondence Address:

TextRelease Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
Tel +31-(0) 20-331.2420 ● info@textrelease.com ● www.textrelease.com

Index to Authors

Forthcoming
February 2019

'Research Data Fuels and Sustains Grey Literature'

Loyola University New Orleans, USA • December 3-4, 2018

Publication Order Form

TWENTIETH INTERNATIONAL CONFERENCE ON GREY LITERATURE

Publication(s):	*No. of Copies*	*x*	*Amount in Euros*	Subtotal
GL20 CONFERENCE PROCEEDINGS – Printed Edition ISBN 978-90-77484-33-3 ISSN 1386-2316 ***Postage and Handling excluded****		x	109.00 =	€
GL20 CONFERENCE PROCEEDINGS - PDF Edition ISBN 978-90-77484-33-3 ISSN 1386-2316 ***Forwarded via email***		x	109.00 =	€
GL20 Conference Proceedings – Online Edition ISBN 978-90-77484-33-3 ISSN 2211-7199 ***Password Protected Access***		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 "GL20 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

GL20 Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
Tel. +31-(0) 20-331.2420 Email: info@textrelease.com